

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2003, 05:56:31 / Search time 8288 Seconds
(without alignments)
11219.544 Million cell updates/sec

Title: US-09-938-803-26

Perfect score: 2273

Sequence: 1 999999tgaagcgatagctt.....ttttcaaaaaaaaaaaaaa 2273

Scoring table: OLIGO NUC
Gapop 60.0, Gapext 60.0

Searched: 2888711 seqs, 2045481386 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_srf:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_srf:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_rod:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2273	100.0	2273	6 AX048067	AX048067 Sequence
2	2211	97.3	2264	6 AX073570	AX073570 Sequence
3	1996	87.8	2058	6 BC000664	BC000664 Homo sapi
4	1868	82.2	2482	6 AK096775	AK096775 Homo sapi
5	1842	81.0	1842	6 AX331926	AX331926 Sequence
6	1842	81.0	1842	6 AX410769	AX410769 Sequence
7	1842	81.0	1842	6 HSU72515	HSU72515 Human C3f
8	1461	64.3	1461	6 AX073560	AX073560 Sequence
9	1258	54.5	1573	6 AK058063	AK058063 Homo sapi
10	1094	48.1	1146	9 BT007000	BT007000 Homo sapi
11	1093	48.1	1146	12 BT007735	BT007735 Synthetic
12	708	31.1	155975	9 AC006512	AC006512 Homo sapi
13	708	31.1	222930	9 HSU47924	HSU47924 Human chrom
14	424	18.7	1641	9 BC009856	BC009856 Homo sapi
15	402	17.7	450	6 BD023986	BD023986 Sequence
16	254	11.2	990	6 AX331531	AX331531 Sequence
17	254	11.2	990	6 AX331927	AX331927 Sequence
18	254	11.2	990	6 AX335065	AX335065 Sequence
19	254	11.2	990	6 HSU72517	HSU72517 Human alter
20	112	4.9	1414	6 AX336765	AX336765 Sequence
21	112	4.9	1414	6 HSU72507	HSU72507 Human 40871
22	73	3.2	280	6 AX409345	AX409345 Sequence
23	64	2.8	120846	9 AL365232	AL365232 Human DNA
24	60	2.6	37223	9 AF135116	AF135116 Homo sapi
25	60	2.6	74245	2 AC022622	AC022622 Homo sapi
26	60	2.6	101805	9 AC073508	AC073508 Homo sapi
27	60	2.6	107532	9 AC010425	AC010425 Homo sapi
28	60	2.6	122900	2 AL139126	AL139126 Homo sapi
29	60	2.6	125527	9 AL353133	AL353133 Human DNA
30	60	2.6	132557	9 AC004520	AC004520 Homo sapi
31	60	2.6	137830	9 AC003029	AC003029 Homo sapi
32	60	2.6	141923	2 AC109997	AC109997 Homo sapi
33	60	2.6	145887	9 AC091435	AC091435 Homo sapi
34	60	2.6	149089	9 AC090735	AC090735 Homo sapi
35	60	2.6	154381	2 AC009096	AC009096 Homo sapi
36	60	2.6	156300	9 AC005919	AC005919 Homo sapi
37	60	2.6	158832	2 AC015475	AC015475 Homo sapi
38	60	2.6	160597	9 AL590666	AL590666 Human DNA
39	60	2.6	168137	9 AC090283	AC090283 Homo sapi
40	60	2.6	175088	9 AC121766	AC121766 Homo sapi
41	60	2.6	176845	9 AC040163	AC040163 Homo sapi
42	60	2.6	181188	2 AC067983	AC067983 Homo sapi
43	60	2.6	184133	9 AC103559	AC103559 Homo sapi
44	60	2.6	194385	9 CDS00001	AL049776 Human chr
45	60	2.6	200807	9 AC073278	AC073278 Homo sapi

ALIGNMENTS

RESULT 1	AX048067	2273 bp	DNA	linear	PAT 15-DEC-2000
LOCUS	AX048067				
DEFINITION	Sequence 61 from Patent WO0070047.				
ACCESSION	AX048067				
VERSION	AX048067.1	GI:11876890			
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens (human)				
	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
REFERENCE					
AUTHORS	Yue, H., Tang, Y.T., Lal, P., Reddy, R., Batra, S., Baughn, M.R.,				
	Yang, J., Azimzai, Y., Lu, D.A., Au-Young, J., and Shih, L.L.				
TITLE	Full-length molecules expressed in human tissues				

Prod. No. is the number of results predicted by chance to have a

JOURNAL Patent: WO 0070047-A 61.23-NOV-2000;
Incyte Genomics, Inc. (US)
Location/Qualifiers
source 1..2273
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 1928920CB1"
BASE COUNT 494 a 630 c 547 g 602 t
ORIGIN
Query Match 100.0%; Score 2273; DB 6; Length 2273;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGGGTGAAGCGATACCTTTTGGCCCGCATTCGGGGCGCGGACCTGGGGGGGTCCTCTG 60
Db 1 GGGGGTGAAGCGATACCTTTTGGCCCGCATTCGGGGCGCGGACCTGGGGGGGTCCTCTG 60
QY 61 GGGCTCCCGAGTTAAGATGGCGTCTCAGCGAGGGGGAGCAGGGGACTGTGTGGCGC 120
Db 61 GGGCTCCCGAGTTAAGATGGCGTCTCAGCGAGGGGGAGCAGGGGACTGTGTGTGGCGC 120
QY 121 TGGGGGGGGTTTGCAGTGGGGTTTCCAGAGGCTGAGCTTTAAACAAGTTGGCGACGTC 180
Db 121 TGGGGGGGGTTTGCAGTGGGGTTTCCAGAGGCTGAGCTTTAAACAAGTTGGCGACGTC 180
QY 181 TGGGGCGGTGAGAACAGGCGCTGGGCGGTGATCATCTTCATCTTCTGGGTTACCCCTTG 240
Db 181 TGGGGCGGTGAGAACAGGCGCTGGGCGGTGATCATCTTCATCTTCTGGGTTACCCCTTG 240
QY 241 CTTTGTGTTATGGGATTAACCTTTTCTCAAGAGACCTTACCTACCTCTTCCATA 300
Db 241 CTTTGTGTTATGGGATTAACCTTTTCTCAAGAGACCTTACCTACCTCTTCCATA 300
QY 301 CCTTTACAGGCTCTCAATGCTTATTTTAACTTTGAAACAGCTTACACCTCCCTGC 360
Db 301 CCTTTACAGGCTCTCAATGCTTATTTTAACTTTGAAACAGCTTACACCTCCCTGC 360
QY 361 TGTGTATTTGCTTCAGTTCCTCAATCTTGCATTAATGGGCGGACCATCAGCGCTCC 420
Db 361 TGTGTATTTGCTTCAGTTCCTCAATCTTGCATTAATGGGCGGACCATCAGCGCTCC 420
QY 421 TCACTACCTTTTGTCCCAATGGCCCTTCTGGCTGGATCTATTAACCTGCCACCG 480
Db 421 TCACTACCTTTTGTCCCAATGGCCCTTCTGGCTGGATCTATTAACCTGCCACCG 480
QY 481 GCAACTACGATATCAATGAGCAATGCAATGCTGTCTGAATTTGAAGCTGATGGT 540
Db 481 GCAACTACGATATCAATGAGCAATGCAATGCTGTCTGAATTTGAAGCTGATGGT 540
QY 541 TGGCTGTGAATACTTTGACGAGGAGAAAGATCAAAATTCCTTGTCTCTGAGCAACA 600
Db 541 TGGCTGTGAATACTTTGACGAGGAGAAAGATCAAAATTCCTTGTCTCTGAGCAACA 600
QY 601 AATATGCGAATACGATGCTTCTCTCCCTGCTGGAAGTTGCTGTTCTCTACTTATG 660
Db 601 AATATGCGAATACGATGCTTCTCTCCCTGCTGGAAGTTGCTGTTCTCTACTTATG 660
QY 661 GGGGCTTCTTGTGAGGCGCCAGTTCTCAATGAATCACTAATGAGCTGAGGGAG 720
Db 661 GGGGCTTCTTGTGAGGCGCCAGTTCTCAATGAATCACTAATGAGCTGAGGGAG 720
QY 721 AGCTGATTTGACATACAGAAAGATACCAACAGCATCATTCCTGCTTCAAGCGCTGA 780
Db 721 AGCTGATTTGACATACAGAAAGATACCAACAGCATCATTCCTGCTTCAAGCGCTGA 780
QY 781 GTCTGGGCTTTTCTACTAGTGGCTACACAGCTCAGCGCCCAATACAGAAACT 840
Db 781 GTCTGGGCTTTTCTACTAGTGGCTACACAGCTCAGCGCCCAATACAGAAACT 840
QY 841 ATCTCTCACTGAAGACTATGACAAACCCCTTCTGTTCCGCTGATGATCACTGCTGA 900
Db 841 ATCTCTCACTGAAGACTATGACAAACCCCTTCTGTTCCGCTGATGATCACTGCTGA 900

Db 841 ATCTCTCACTGAAGACTATGACAAACCCCTTCTGTTCCGCTGATGATCACTGCTGA 900
QY 901 TCTGGGGCAAGTTTGTCTGTACAAATATGTCACTTTGGCTGTGACAGAAAGATAT 960
Db 901 TCTGGGGCAAGTTTGTCTGTACAAATATGTCACTTTGGCTGTGACAGAAAGATAT 960
QY 961 GCATTTTGAACGGGCTTCAATGCTTTGAAAGAAAAGGCAAGCAAGCTGGAGT 1020
Db 961 GCATTTTGAACGGGCTTCAATGCTTTGAAAGAAAAGGCAAGCAAGCTGGAGT 1020
QY 1021 CCTGTGCCAAATGAAGTGTGGCTTTTGAACCAACCCCGCTTCACTGAGCAATTTG 1080
Db 1021 CCTGTGCCAAATGAAGTGTGGCTTTTGAACCAACCCCGCTTCACTGAGCAATTTG 1080
QY 1081 CCTGATTTCAACATCAACCAACGCTGGGCTGCGCTCAATCTTCAACGACTCAAGT 1140
Db 1081 CCTGATTTCAACATCAACCAACGCTGGGCTGCGCTCAATCTTCAACGACTCAAGT 1140
QY 1141 TCTTGGAAATTAAGAACTCTCAGAGTCTCTGTTGCTATTCTCGGCTCTGGCAGC 1200
Db 1141 TCTTGGAAATTAAGAACTCTCAGAGTCTCTGTTGCTATTCTCGGCTCTGGCAGC 1200
QY 1201 GCTGCACTCAGGATACCTGTCTGCTTCAGATGGAATTCATTTGTTGTTGTTGTTG 1260
Db 1201 GCTGCACTCAGGATACCTGTCTGCTTCAGATGGAATTCATTTGTTGTTGTTGTTG 1260
QY 1261 GACAGGCTGCAAGCTCATTTCAAGAGAGCCCACTGAGCAAGCTGGCCGCTTATCTG 1320
Db 1261 GACAGGCTGCAAGGCTCATTTCAAGAGAGCCCACTGAGCAAGCTGGCCGCTTATCTG 1320
QY 1321 TCTTCAAGCCCTTCTAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1380
Db 1321 TCTTCAAGCCCTTCTAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1380
QY 1381 CCATGACTGCT 1440
Db 1381 CCATGACTGCT 1440
QY 1441 ATTTCTTGGCCCAATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1500
Db 1441 ATTTCTTGGCCCAATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1500
QY 1501 CAATGTTGCCAAGGAAAGGAAATTAAGAGTGAATGATTCATTTCCCTGCTGCTCT 1560
Db 1501 CAATGTTGCCAAGGAAAGGAAATTAAGAGTGAATGATTCATTTCCCTGCTGCTCT 1560
QY 1561 GTGCGGAGCTGTGAGAAACTACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
Db 1561 GTGCGGAGCTGTGAGAAACTACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
QY 1621 CAGAGATGGAAGAGCCAGGAGCTGAAAGATGATGCTTCCAGCTGTGCTCTGCTCC 1680
Db 1621 CAGAGATGGAAGAGCCAGGAGCTGAAAGATGATGCTTCCAGCTGTGCTCTGCTCC 1680
QY 1681 AGCCAGTCTTCAATTTGGGGCCAAAGGGGAACTTTTGTGGAAGAGGCTGTGCTT 1740
Db 1681 AGCCAGTCTTCAATTTGGGGCCAAAGGGGAACTTTTGTGGAAGAGGCTGTGCTT 1740
QY 1741 GTCAACCAACGCTGGAATGCAAGTGGCGGATCTCAGCTCAACCGAACCCTCCACTCT 1800
Db 1741 GTCAACCAACGCTGGAATGCAAGTGGCGGATCTCAGCTCAACCGAACCCTCCACTCT 1800
QY 1801 TTCAAGTATTTTCTGCTCTCAGCTTCCCAAGTACGCTGGGAATACAGGCAAGCAATG 1860
Db 1801 TTCAAGTATTTTCTGCTCTCAGCTTCCCAAGTACGCTGGGAATACAGGCAAGCAATG 1860
QY 1861 CCCAGCTAATTTTGTATTTTGTAGTGAAGAGGGAATTTCAAGCTTTGGCAGGCTGCT 1920
Db 1861 CCCAGCTAATTTTGTATTTTGTAGTGAAGAGGGAATTTCAAGCTTTGGCAGGCTGCT 1920
QY 1921 TCGAATCTCTGACCGCAAGTATCAACCGCTCTCGGCTTCCAAAGTGTGTTGTTTCAAG 1980
Db 1921 TCGAATCTCTGACCGCAAGTATCAACCGCTCTCGGCTTCCAAAGTGTGTTGTTTCAAG 1980

OY		1961	GCGTGAAGCCACCGGTGCGGCCGCCCAAAAGGGAAAACTTTGTGGAGAGCAGAGGGGGCTCA	2040
Db		1961	GCGTGAAGCCACCGGTGCGGCCGCCCAAAAGGGAAAACTTTGTGGAGAGCAGAGGGGGCTCA	2040
OY		2041	CATCGCCCCCTGMAATTCCCAGATGACAATTGCCTTAATCTGCCCATTTACCAGAGATC	2100
Db		2041	CATCTCCCCCTTGATTTCCCCAGATGACAATTGCTTAATCTGCCCATTTACCAGAGATC	2100
OY		2101	TATGTGTGTTTTCTTCCTGCGCAATTACTCATGATTGTGTATGTGTCGCGTAGCACACCCCC	2160
Db		2101	TATGTGTGTTTTCTTCCTGCGCAATTACTCATGATTGTGTATGTGTCGCGTAGCACACCCCC	2160
OY		2161	CCCATGGGGGGGTGAGAGAAGGGGTGCAAGCCCTGCTGCTCCACTTTTCTTACCTTGGA	2220
Db		2161	CCCATGGGGGGGTGAGAGAAGGGGTGCAAGCCCTGCTGCTCCACTTTTCTTACCTTGGA	2220
OY		2221	CTGTATTGATAAATCACTTCTGTGTTGTTCAGTTTTCAAAIAAAAAAAAAAAAAA	2273
Db		2221	CTGTATTGATAAATCACTTCTGTGTTGTTCAGTTTTCAAAIAAAAAAAAAAAAAA	2273
RESULT 2				
LOCUS	AX073570	2264 bp	DNA	linear
DEFINITION	Sequence 22 from Patent WO0104297.			PAT 06-FEB-2001
ACCESSION	AX073570			
VERSION	AX073570.1	GI:12709984		
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	Kato,S. and Kimura,T.			
AUTHORS	1			
TITLE	Human proteins having hydrophobic domains and dna encoding these			
JOURNAL	Patent: WO 0104297-A 22 18-JAN-2001;			
FEATRES	SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Proteogene Inc. (JP)			
	Location/Qualifiers			
source	1..2264			
	/organism="Homo sapiens"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:9606"			
	85..11548			
	/note="unnamed protein product"			
	/codon_start=1			
	/protein_id="CAC28402.1"			
	/db_xref="GI:12709985"			
	/translation="MASASBEGDEGVVALAGVLSQGEIISLANKATSLIASRQAALILISFLGPALFYRHLYPKRETYLIHLFHFPTGLSIALYPRFGNQLYHSLSICTYLQTLIRLMGRITTVLTTPCFOMAYVLIAGYITRTNGVDIKMTNPICVLLKILGLAVDFDGGSDNSLEEQKRYAIRVPVSLLEVAFSEYFYDGALVVGQFSNMHYMKLVQDELIDIPGKI PMSII PALRPLSGFLYLVGYTLSPHILEDYLDIEDYNWHPFWFCMYMLTIWGKPYLVXVCMLVTIEGVCLITGLGFNGFBEKGAKWDACAMKMYLFETNPRTGTIGASFNNTAWVAWARYIFYEKLKFTGNKESSQSLFLPILMHGLSHGYLVCPQMEELIYIVEROARALIOESPITSKLKAITVLAOPFYLVQOITIMWLPMQYSMTACLTPWDMKLAVYKSYVLCIHIFLPSLPIPLFIYHHAMPKKELKKME"			
CDS				
BASE COUNT	482 a	629 c	550 g	603 t
ORIGIN				
Query Match	97.3%; Score 2211; DB 6; Length 2264;			
Best Local Similarity	100.0%; Pred. No. 0;			
Matches 2211; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
OY		47	GGGGGGTCCCTGTGTGGGGCTCCCGAGTTAAAGATGCGTCTCAGCGAGGGGAGCAGAGG	106
Db		54	GGGGGGTCCCTGTGTGGGGCTCCCGAGTTAAAGATGCGTCTCAGCGAGGGGAGCAGAGG	113
OY		107	GACTGTGTGTGGCGCTGGGGGGGGTCTGTGCACTCGCGTTTCCAGAGACTGACCTTTAACAA	166
Db		114	GACTGTGTGTGGCGCTGGGGGGGGTCTGTGCACTCGCGTTTCCAGAGACTGACCTTTAACAA	173

QY	167	GTTGGGAGAGCTCCCTGGGGCGGTGAAACAGGCGCTGGGGCTGATCATCTCATCTTCCCT	226
Db	174	GTTGGCGAGCTCCCTGGGGCGGTGAAACAGGCGCTGGGGCTGATCATCTCATCTTCCCT	233
QY	227	GGGTAAACCCCTTGGCTTGTATTATTCGGGCAATTAACCTTTTCTACAGAGAACCTTCAAT	286
Db	234	GGGTAAACCCCTTGGCTTGTATTATTCGGGCAATTAACCTTTTCTACAGAGAACCTTCAAT	293
QY	287	CCACCTCTTCATATACCTTTTACAGGCGCTCTCAATTGCTTAATTTTAATTTTGAACCCAGCT	346
Db	294	CCACCTCTTCATATACCTTTTACAGGCGCTCTCAATTGCTTAATTTTAATTTTGAACCCAGCT	353
QY	347	CTACCACTCTCCCTGCTGTGTATTTGTGCTTCAGTTTCTCATCTTTCGATATATGGGCGGAC	406
Db	354	CTACCACTCTCCCTGCTGTGTATTTGTGCTTCAGTTTCTCATCTTTCGATATATGGGCGGAC	413
QY	407	CATCACTGCGGCTCCTACATACCTTTTGGCTTCCAGATAGGCGCTCAACCTTCTGAGCTGGAATCA	466
Db	414	CATCACTGCGGCTCCTACATACCTTTTGGCTTCCAGATAGGCGCTCAACCTTCTGAGCTGGAATCA	473
QY	467	TTACACTGCGACCGGCAACTACGATATCAAGTGGACAATGCCACATTGTTCGACTTT	526
Db	474	TTACACTGCGACCGGCAACTACGATATCAAGTGGACAATGCCACATTGTTCGACTTT	533
QY	527	GAAGCTGATTTGGTTTGGCTGTGTACTTCTTAACGAGGGAAAGATCAGAATTCCTTGTTC	586
Db	534	GAAGCTGATTTGGTTTGGCTGTGTACTTCTTAACGAGGGAAAGATCAGAATTCCTTGTTC	593
QY	587	CTCTAGAGCAACGAATATATGCCATACGTGTGTCTTCCCTCCGCTGGAGTTGTCTGCTTT	646
Db	594	CTCTAGAGCAACGAATATATGCCATACGTGTGTCTTCCCTCCGCTGGAGTTGTCTGCTTT	653
QY	647	CTCCTACTTCTATGGGGGCTTCTTGGTAGGGGCCCGCAGTTCTCAATGAAATCACTACATGAA	706
Db	654	CTCCTACTTCTATGGGGGCTTCTTGGTAGGGGCCCGCAGTTCTCAATGAAATCACTACATGAA	713
QY	707	GCTGTGCAAGGAGAGCTGATTTGACATACAGAGAAAGATACCAACAGCATCATTTCTGCTC	766
Db	714	GCTGTGCAAGGAGAGCTGATTTGACATACAGAGAAAGATACCAACAGCATCATTTCTGCTC	773
QY	767	TCTCAAGGGCCCTGAATCTGGGCTTTTCTACCTATGTGGCTTACACCTGCTCAGCCCCCA	826
Db	774	TCTCAAGGGCCCTGAATCTGGGCTTTTCTACCTATGTGGCTTACACCTGCTCAGCCCCCA	833
QY	827	CATCAACAGAAAGCTATCTCTCTCACTGAAGACTATGACAAACACCCCTTCTGGTTCCGCTG	886
Db	834	CATCAACAGAAAGCTATCTCTCTCACTGAAGACTATGACAAACACCCCTTCTGGTTCCGCTG	893
QY	887	CATGTACATGCTGATCTGGGGCAAGTTGTGTGTCGAATAATGTCAACCTGTGTGGCTGTGT	946
Db	894	CATGTACATGCTGATCTGGGGCAAGTTGTGTGTCGAATAATGTCAACCTGTGTGGCTGTGT	953
QY	947	CACAGAAAGAGATATGCAATTTTGAACGGGCTTGGGCTTCAATGCGCTTTGAAGAAAGGGCAA	100
Db	954	CACAGAAAGAGATATGCAATTTTGAACGGGCTTGGGCTTCAATGCGCTTTGAAGAAAGGGCAA	101
QY	1007	GGCAAAAGTGGAGTGCCTGTGCGCAACATGAAGGTGTGGCTCTTTGAAACAAACCCCGGCTT	106
Db	1014	GGCAAAAGTGGAGTGCCTGTGCGCAACATGAAGGTGTGGCTCTTTGAAACAAACCCCGGCTT	107
QY	1067	CACGTGGCACATTTGCTCATTCACATCAACCAACGCGCTGGGTGGCCCGCTACATCTTT	112
Db	1074	CACGTGGCACATTTGCTCATTCACATCAACCAACGCGCTGGGTGGCCCGCTACATCTTT	113
QY	1127	CAAAAGCATCAAGTTCTTTGGAAATTAAGAACTCTCTCAGGGTCTCTGTTGCTATTTCTT	118
Db	1134	CAAAAGCATCAAGTTCTTTGGAAATTAAGAACTCTCTCAGGGTCTCTGTTGCTATTTCTT	119
QY	1187	GGCCCTCTGGGACGGGCTGCACTAGAGATACCTGTGTGTCTTCCAGATGGAAATTCCTCAT	124
Db	1194	GGCCCTCTGGGACGGGCTGCACTAGAGATACCTGTGTGTCTTCCAGATGGAAATTCCTCAT	125
QY	1247	TGTTATTTGTGAATAACAGGCTGCCAGGCTCATTTCAAGAGAGCCCAACCTCTAGCAAGCT	130

Db 1254 TGTATTGGAGAAAGACAGCTGCCAGCTCATTCAGAGAGCCCACTTGAGCAAGCT 1313
 QY 1307 GGGCGGCATTAATGCTCTCCAGCCCTTCTACTATTTGGTGCACAGACCATTCACAGGCT 1366
 Db 1314 GGGCGGCATTAATGCTCTCCAGCCCTTCTACTATTTGGTGCACAGACCATTCACAGGCT 1373
 QY 1367 CTTTCATGGGTTACTCATGACGCTCTTGCCTTTACGCTGAGCAAAATGCTTAAGT 1426
 Db 1374 CTTTCATGGGTTACTCATGACGCTCTTGCCTTTACGCTGAGCAAAATGCTTAAGT 1433
 QY 1427 GTTAAATCCATCTATTTCTGTCGACCATCTTCTGTCGACCTATCTATCTATTTGCC 1486
 Db 1434 GTTAAATCCATCTATTTCTGTCGACCATCTTCTGTCGACCTATCTATCTATTTGCC 1493
 QY 1487 TTATATTCACAAAGCAATGCTGTCAGAGAAAGAAAGTTAAAGATGATATTCAT 1546
 Db 1494 TTATATTCACAAAGCAATGCTGTCAGAGAAAGAAAGTTAAAGATGATATTCAT 1553
 QY 1547 TTCCTGTGGCTGTGCGGAGCTGTGCAAGAACTACTGCTCTCCCTTTACAGCACT 1606
 Db 1554 TTCCTGTGGCTGTGCGGAGCTGTGCAAGAACTACTGCTCTCCCTTTACAGCACT 1613
 QY 1607 CTTTGGCCCAAGAGAGAGAAATGGAAGAGCAGGAGGTGGAAGATGATGCTTCCAGCT 1666
 Db 1614 CTTTGGCCCAAGAGAGAGAAATGGAAGAGCAGGAGGTGGAAGATGATGCTTCCAGCT 1673
 QY 1667 GTGCTCTGTGCTGACAGCAAGCTCTTCAATTTGGGCGCAAGAGGAACTTTTGGAGA 1726
 Db 1674 GTGCTCTGTGCTGACAGCAAGCTCTTCAATTTGGGCGCAAGAGGAACTTTTGGAGA 1733
 QY 1727 AGGCGTCTGCTTTGTTCACCCAGCGCTGGAATGCAAGTGGCGGATCTCAGCTCACCGCAAC 1786
 Db 1734 AGGCGTCTGCTTTGTTCACCCAGCGCTGGAATGCAAGTGGCGGATCTCAGCTCACCGCAAC 1793
 QY 1787 CTCACCTCTGCTGAGTCAAGTGAATTTTCGCGCTCAGCCCTCCCAAGTATGCTGGGAATACA 1846
 Db 1794 CTCACCTCTGCTGAGTCAAGTGAATTTTCGCGCTCAGCCCTCCCAAGTATGCTGGGAATACA 1853
 QY 1847 GGCACGCGCACCATGCCAGCTAATTTTGTATTTTCAGTAAACGGGATTTTCCACAGCT 1906
 Db 1854 GGCACGCGCACCATGCCAGCTAATTTTGTATTTTCAGTAAACGGGATTTTCCACAGCT 1913
 QY 1907 TGGCCAGGCTGTGCTGCACTCTGACCGCAAGTATCCACCGCTCCCTCCCAAG 1966
 Db 1914 TGGCCAGGCTGTGCTGCACTCTGACCGCAAGTATCCACCGCTCCCTCCCAAG 1973
 QY 1967 TGGCGGATTAACAGGCGTGAAGCAAGTGGCCCGGCCCAAGGGGAACTCTTGGGAGG 2026
 Db 1974 TGGCGGATTAACAGGCGTGAAGCAAGTGGCCCGGCCCAAGGGGAACTCTTGGGAGG 2033
 QY 2027 AGCAGAGGGCTCAATCTCCCTCTGATTTCCCATGACATTTGCTTCTCCCA 2086
 Db 2034 AGCAGAGGGCTCAATCTCCCTCTGATTTCCCATGACATTTGCTTCTCCCA 2093
 QY 2087 TCTAGCCAGAAATCTATTTGTTTCTTCTGCGCAATTTACTATGATTTGATGTGCCG 2146
 Db 2094 TCTAGCCAGAAATCTATTTGTTTCTTCTGCGCAATTTACTATGATTTGATGTGCCG 2153
 QY 2147 CTACCCACACCCCCCATGGGGGGGTGAGAGGGGTGCAAGGCGCTGCTCCCACTT 2206
 Db 2154 CTACCCACACCCCCCATGGGGGGGTGAGAGGGGTGCAAGGCGCTGCTCCCACTT 2213
 QY 2207 TTCTTACCTTGAAGTCTATTTAGTAAATCACTTCTGTTGTTCAGTTT 2257
 Db 2214 TTCTTACCTTGAAGTCTATTTAGTAAATCACTTCTGTTGTTCAGTTT 2264

RESULT 3
 BC000664
 LOCUS
 DEFINITION
 (CDNA clone MGC:1311 IMAGE:3349388), complete cds.

ACCESSION BC000664
 VERSION BC000664.1 GI:12653756
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2058)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Mang, J., Heich, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schein, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, X.S., Krzywinski, M.I., Skalske, U., Small, D.E., Scherch, A., Schein, U.E., Jones, S.J. and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 2 (bases 1 to 2058)
 Strausberg, R.
 Direct Submision
 Submitted (15-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 CONTACT
 Contact: MGC help desk
 Email: cgabs-room1.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakeley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Khong, P., Latic, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantirpop, S., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 5 Row: d Column: 14
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 15375317.
 Location/Qualifiers
 1. 2058
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:1311 IMAGE:3349388"
 /tissue_type="Colon, adenocarcinoma"
 /clone_lib="NIH MGC 15"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"

gene 1. .2058
/gene="C3F"
/db_xref="LocusID:10162"
177. .1322
/codon_start=1
/product="putative protein similar to nesey"
/protein_id="AAH0664.1"
/db_xref="GI:12653757"
/db_xref="LocusID:10162"
/translation="MGRTITAVLTFCFQMAVLLAGYYTAATGNVDIKWMPHCYVLT.
MLUGVELIDIPKIPNSITPAKRLSLGIFLVGTLLSPHITLEDYLDNHP
WRCKMILMGKFLVLYKVTCLWLTGVCILTGLGNGEKEKAKMDACANNKWL
ETNPRTGTIASFNINAMAVARYIFKRLFKLNGKLSQSLSLFALMHGHSGLV
CEQMKELIVIVERQARLIQESPTLSKLAITVLQPFYLVQOTIHLFMGVSMTAF
LFTWDMKLKYSIYFLGHLFSLFLPIYIKHAWPKKKKKME"

BASE COUNT 470 a 575 c 457 g 556 t

ORIGIN

Query Match 87.8%; Score 1996; DB 9; Length 2058;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2046; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 227 GGGTTACCCCTTGGCTTTGTTTATGGCATTCCTTTCTACAGAGACTCACTCAT 286
DB 8 GGGTTACCCCTTGGCTTTGTTTATGGCATTCCTTTCTACAGAGACTCACTCAT 67

QY 287 CCACCTCTTCACATACCTTTACAGGGCTCTCAATGCTTTATTTTACCTTTGGAAACAGCT 346
DB 68 CCACCTCTTCACATACCTTTACAGGGCTCTCAATGCTTTATTTTACCTTTGGAAACAGCT 127

QY 347 CTACCACTCCCTGCTGTGATGCTTGAATTCCTCATCTTGAATGAGGCGGCAC 406
DB 128 CTACCACTCCCTGCTGTGATGCTTGAATTCCTCATCTTGAATGAGGCGGCAC 187

QY 407 CATCACTGCGTCTCTACTACCTTTGCTTCCAGATGCGCTTCTTGGCTGATATCTA 466
DB 188 CATCACTGCGTCTCTACTACCTTTGCTTCCAGATGCGCTTCTTGGCTGATATCTA 247

QY 467 TTACACTGCGACCGGCACTAGATATCAAGTGAATGAGCAATGCTTGTCTGACTT 526
DB 248 TTACACTGCGACCGGCACTAGATATCAAGTGAATGAGCAATGCTTGTCTGACTT 307

QY 527 GAAGCTGATGTTGGCTGTGACTACTTCTTGAACGAGGAGAAAGATCAGAAATTCCTTGT 586
DB 308 GAAGCTGATGTTGGCTGTGACTACTTCTTGAACGAGGAGAAAGATCAGAAATTCCTTGT 367

QY 587 CTCTGAGCAACGAATATGCGATACGCTGCTTCTTCCCTGCTGGAAGTGTGCTGTT 646
DB 368 CTCTGAGCAACGAATATGCGATACGCTGCTTCTTCCCTGCTGGAAGTGTGCTGTT 427

QY 647 CTCTACTCTATGAGGCGCTTCTGCTAGAGGCGCCAGCTCTCAATGAATACATCACTGAA 706
DB 428 CTCTACTCTATGAGGCGCTTCTGCTAGAGGCGCCAGCTCTCAATGAATACATCACTGAA 487

QY 707 GCTGTGACGAGGAGAGCTGATGACATACACAGAAAGATACCAACAGCATCTTCTGCT 766
DB 488 GCTGTGACGAGGAGAGCTGATGACATACACAGAAAGATACCAACAGCATCTTCTGCT 547

QY 767 TCTCAAGCGCTGAGTCTGAGGCTTTTCTAAGTGGCTACACATGCTCAGCCCCCA 826
DB 548 TCTCAAGCGCTGAGTCTGAGGCTTTTCTAAGTGGCTACACATGCTCAGCCCCCA 607

QY 827 CATCAGAGAGACTATCTCTCAGTGAAGTATGACACACACCCCTTGTGTTCCGCTG 886
DB 608 CATCAGAGAGACTATCTCTCAGTGAAGTATGACACACACCCCTTGTGTTCCGCTG 667

QY 887 CATGATACATGCTGATCTGGGCAAGTTTGTGCTGTACAAATATGTCACTGTTGGCTGT 946
DB 668 CATGATACATGCTGATCTGGGCAAGTTTGTGCTGTACAAATATGTCACTGTTGGCTGT 727

QY 947 CACAGAGAGATGATGATTTTGAACGAGGCTGAGCTTCAATGCTTTGAAGAAAGGCA 1006

DB 728 CACAGAGAGATGATGATTTTGAACGAGGCTGAGCTTCAATGCTTTGAAGAAAGGCA 787

QY 1007 GGCAGAGTGGATGCTCTGTGCGCAACTGAAGTGTGCTCTTTGAAGAAACCCCGCTT 1066

DB 788 GGCAGAGTGGATGCTCTGTGCGCAACTGAAGTGTGCTCTTTGAAGAAACCCCGCTT 847

QY 1067 CACTGACACATGCTCTCAATTCATCAACACCAAGCGCTGGGTGGCGGCTACATCTT 1126

DB 848 CACTGACACATGCTCTCAATTCATCAACACCAAGCGCTGGGTGGCGGCTACATCTT 907

QY 1127 CAAAGACTCAAGTCTCTTGAATTAAGAACTCTCAAGGCTCTGCTGCTATTCCT 1186

DB 908 CAAAGACTCAAGTCTCTTGAATTAAGAACTCTCAAGGCTCTGCTGCTATTCCT 967

QY 1187 GGCCTCTGCGACGCTCTGCACTCAGAGTATCTGCTCTTCCATGGAATTCCTCAT 1246

DB 968 GGCCTCTGCGACGCTCTGCACTCAGAGTATCTGCTCTTCCATGGAATTCCTCAT 1027

QY 1247 TGTATTTGGAAGAACAGGCTGCGAGCTCATTTCAAGAGAGCCCACTGAGCAAGCT 1306

DB 1028 TGTATTTGGAAGAACAGGCTGCGAGCTCATTTCAAGAGAGCCCACTGAGCAAGCT 1087

QY 1307 GGCCTCATTAAGTCTCTGCTGCGAGCTTCTAATTTGTGCAACAGACCATCAGCT 1366

DB 1088 GGCCTCATTAAGTCTCTGCTGCGAGCTTCTAATTTGTGCAACAGACCATCAGCT 1147

QY 1367 CTTCATGAGTTACTCCATACATGCTCTTCTGCTCTTCAAGTGGACAAATGCTTAAGT 1426

DB 1148 CTTCATGAGTTACTCCATACATGCTCTTCTGCTCTTCAAGTGGACAAATGCTTAAGT 1207

QY 1427 GTATAAATCATATTAATTTCTTGGGACACATCTTCTTCTGAGGCTTCAATATATGCT 1486

DB 1208 GTATAAATCATATTAATTTCTTGGGACACATCTTCTTCTGAGGCTTCAATATATGCT 1267

QY 1487 TTATATTCACAAAGCAATGCTGCCAAGAAAGAAAGAAATTAAGATGAATATCAT 1546

DB 1268 TTATATTCACAAAGCAATGCTGCCAAGAAAGAAAGAAATTAAGATGAATATCAT 1327

QY 1547 TTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1606

DB 1328 TTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1387

QY 1607 CCTTTGCCCAAGAGAGAGAAATGAGAAACCAAGGAGTGAAGATCGATGCTTCCAGCT 1666

DB 1388 CCTTTGCCCAAGAGAGAGAAATGAGAAACCAAGGAGTGAAGATCGATGCTTCCAGCT 1447

QY 1667 GTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1726

DB 1448 GTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1507

QY 1727 AGGCGTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1786

DB 1508 AGGCGTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1567

QY 1787 CTCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1846

DB 1568 CTCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1627

QY 1847 GGCAGGCCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1906

DB 1628 GGCAGGCCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1687

QY 1907 TGGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1966

DB 1688 TGGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1747

QY 1967 TGTGGGATTAAGGCTGAGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2026

DB 1748 TGTGGGATTAAGGCTGAGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1807

QY 2027 AGCAGAGGAGCTCACATCTCCCTCTGATTTCCCTCATGCACTTCTATCTTCCCA 2086

DB 1808 AGCAGAGGAGCTCACATCTCCCTCTGATTTCCCTCATGCACTTCTATCTTCCCA 1867

Oy		2087	TCTGACCGAGAAATCATATTGTTCCTTTTCTTGCCCAATTACTGAATGGTAATGC	2146	
Dd		1668	TCTGACCAGAATACTAATTGTGTTTTTCTCCTGCCAATTTACTAGATTTGTAATGCC	1927	
Oy		2147	CTACCAcAcaccccccccatggsgggttgagagaggcgatcaaggcccctgcctcacatt	2206	
Dd		1928	CTaccaccaCCCCCCCcatggggggttgaAgagggttCaAGgcccttcctCcactt	1987	
Oy		2207	TTTCTACCTTGGAActgatTaGaTTaaAttaCaTTCtTGTTCAGTTTTCaaaaaaa	2266	
Dd		1988	TTTCTACCTTGGAActgatTaGaTTaaAttaCaTTCtTGTTCAGTTTTCaaaaaaa	2047	
Oy		2267	AAAAAAA 2273 2048 AAAAAAAAA 2054		
RESULT_4					
AKO96775					
LOCUS					
DEFINITION	Homo sapiens CDNA FLJ39456 fis, clone PROST2010782, highly similar to Human C3f mRNA.	2482 bp	mRNA linear	PRI 15-JUN-2002	
ACCESSION	AKO96775				
VERSION	AKO96775.1 GI:21756342				
KEYWORDS	oligo capping; fis (full insert sequence).				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE					
AUTHORS	Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T., Kodaira,H., Furuya,T., Takahashi M., Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Kaneita,N., Sato,K., Tanikawa,M., Yamazaki.M., Sugiyama.T., Irie,R., Otsuki,T., Sato H., Wakamatsu,A., Ishii,S., Yumoto,C.J., Isono.Y., Kawai-Hio,Y., Saito.K., Nishikawa.T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura.Y., Sekine.M., Kikuchi.H., Murakawa,K., Kanemori,K., Takahashi-Fujii.A., Oshtima,A., Sugiyama.A., Kawachi.B., Suzuki.Y., Sugano.S., Nagahari.K., Masuko.Y., Nagai.K. and Isogai,T. NEBO human cDNA sequencing project Unpublished 2 (bases 1 to 2482) Isogai,T. and Yamamoto,J. Direct Submission Submitted (04-JUN-2002) Takao Isogai, Fuj Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail:genom@esohri.co.jp Tel.:81-438-52-3975 Fax:81-438-52-3986) NEBO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institite of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB. Location/Qualifiers 1..2482 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /cclone="PROST2010782" /tissue.type="prostate" /clone.lib="PROS73" /note="cloning vector: PME18FPL3"				
BASE COUNT	526 a 687 c 615 g 654 t				
ORIGIN					
Query Match	82.2%; Score 1688; DB 9; Length 2482;				
Best Local Similarity	99.8%; Pred. No. 0;				
Matches 2208; Conservative	0; Mismatches 2; Indels 2; Gaps 2				
47 GGGGGGTCCCTGTGGGGCTCCCGAGTTAAGATGCGCTCCTTAGCGGAGGGGACAAGG	106				

Db	273	GGGGGGTTCCTGTGAGGCTCCCGAGTTAAGATGGCTCTCAAGCGAAGGGGACGAGGG	332
Qy	107	GACTGTGTGGCGCTGGCGGGGGGTTCTGCAGTGGGTTTTCCAGAGCTGAGCTTAA	166
Db	333	GACTGTGTGGCGCTGGCGGGGGGTTCTGCAGTGGGTTTTCCAGAGCTGAGCTTAA	392
Qy	167	GTTGGCGACGTCCCTGGGGCGGTACAAACAGCGCTCGGCTGATCATCTCCACTTCT	226
Db	393	GTTGGCGACGTCCCTGGGGCGGTACAAACAGCGCTCGGCTGATCATCTCCACTTCT	452
Qy	227	GGGTTACCCCTTGTCTTGTATTATGCGGCACTTCTTTCTACAGAGACCTTCACT	286
Db	453	GGGTTACCCCTTGTCTTGTATTATGCGGCACTTCTTTCTACAGAGACCTTCACT	512
Qy	287	CCACCTCTTCATACCTTTACAGGCGCTCAATGTCTTATTTTAACTTTGGAAACAGT	346
Db	513	CCACCTCTTCATACCTTTACAGGCGCTCAATGTCTTATTTTAACTTTGGAAACAGT	572
Qy	347	CTACCACTCCCTGCTGTGTATTGTGCTTCACTTCTCATCTTGCATTAATGGCGCGAC	406
Db	573	CTACCACTCCCTGCTGTGTATTGTGCTTCACTTCTCATCTTGCATTAATGGCGCGAC	632
Qy	407	CATCACTGCGGCTCCCTCACTACCTTTGCTTCCAGATGGCGCTTCTTGCGGTGATCTA	466
Db	633	CATCACTGCGGCTCCCTCACTACCTTTGCTTCCAGATGGCGCTTCTTGCGGTGATCTA	692
Qy	467	TTAACACTGCCACCGGCACTACGATATCAAGTGACAATGCGACATGTGTCTGACTTT	526
Db	693	TTAACACTGCCACCGGCACTACGATATCAAGTGACAATGCGACATGTGTCTGACTTT	752
Qy	527	GAAGCTGATGTGTGTGCTGTGTACTTCTTGCACGGAGGAAATGATGAAATTCCTGTC	586
Db	753	GAAGCTGATGTGTGTGCTGTGTACTTCTTGCACGGAGGAAATGATGAAATTCCTGTC	812
Qy	587	CTCTGAGCAACAGAAATATGCCATACGATGGTGTTCTTCCCTGTGGAAGTGTGTTT	646
Db	813	CTCTGAGCAACAGAAATATGCCATACGATGGTGTTCTTCCCTGTGGAAGTGTGTTT	872
Qy	647	CTCTACTTCTATGGGGCTTCTTGGTAGGGGCCCAAGTCTCAATGATCACTACATGAA	706
Db	873	CTCTACTTCTATGGGGCTTCTTGGTAGGGGCCCAAGTCTCAATGATCACTACATGAA	932
Qy	707	GCTGTGACGGGAGACGTATTGACATACAGGAAAGTATCCAAACAGCATCATCTCTGC	766
Db	933	GCTGTGACGGGAGACGTATTGACATACAGGAAAGTATCCAAACAGCATCATCTCTGC	992
Qy	767	TCTCAAGCGCTGAGTCTGGGCTTTTCTACCTTGTGGGCTACACACTGCTCACGCCCA	826
Db	993	TCTCAAGCGCTGAGTCTGGGCTTTTCTACCTTGTGGGCTACACACTGCTCACGCCCA	1052
Qy	827	CATCAACAGACATATCTCTCACTGAGACTATGACAACCAACCCCTTGTGGTCCGCTG	886
Db	1053	CATCAACAGACATATCTCTCACTGAGACTATGACAACCAACCCCTTGTGGTCCGCTG	1112
Qy	887	CATGTACATGCTGATCTGGGGCAAGTTGTGCTGTACAAATATGTCACTTGTGGCTGT	946
Db	1113	CATGTACATGCTGATCTGGGGCAAGTTGTGCTGTACAAATATGTCACTTGTGGCTGT	1172
Qy	947	CACAGAAAGAGATGACTTTTGAAGGGCGCTGGCTCATAGGCTTTGAAAGAAAGGCGAA	1006
Db	1173	CACAGAAAGAGATGACTTTTGAAGGGCGCTGGCTCATAGGCTTTGAAAGAAAGGCGAA	1232
Qy	1007	GGCAAAAGTGGATGCTGTGCAACATGAAGGTGTGCTTTTGAACAAACCCCGCTT	1066
Db	1233	GGCAAAAGTGGATGCTGTGCAACATGAAGGTGTGCTTTTGAACAAACCCCGCTT	1292
Qy	1067	CACGTGGCACTTGTGCTTATTCACATTAACCAAGCCCTGGGTGGCGGCTCATCTT	1126
Db	1293	CACGTGGCACTTGTGCTTATTCACATTCACCAAGCCCTGGGTGGCGGCTCATCTT	1352
Qy	1127	CAAAAGACTCAAGTCTCTGGAATTAAGAACTCTCAGGGGTCTCTGTTGATTAATCT	1186

Db 1353 CAAAGCACTCAAGTCTCTGGAAAT-AAATACTCTTAAAGGCTCTCTGTTGCTATTCT 1411
QY 1187 GGGCCCTCTGGACGGGCTGCACTCAGGAATCTGTCTGCTTCAGATGGAAATTCCTCAT 1246
Db 1412 GGGCCCTCTGGACGGGCTGCACTCAGGAATCTGTCTGCTTCAGATGGAAATTCCTCAT 1471
QY 1247 TGTATTGTGGAAAGACAGGCTGCCAGGCTCATTTCAAGAGAGCCCACTGAGCAAGCT 1306
Db 1472 TGTATTGTGGAAAGACAGGCTGCCAGGCTCATTTCAAGAGAGCCCACTGAGCAAGCT 1531
QY 1307 GGGCCGCACTTACTCTCTCCAGCCCTTCTACTATTGTGTGCAACAACCATCCATGCTGCT 1366
Db 1532 GGGCCGCACTTACTCTCTCCAGCCCTTCTACTATTGTGTGCAACAACCATCCATGCTGCT 1591
QY 1367 CTTCATGGGTTACTCCATGACTGCTCTTCTGCTCTTCAAGTGGGAGCAAAATGGTTAAGGT 1426
Db 1592 CTTCATGGGTTACTCCATGACTGCTCTTCTGCTCTTCAAGTGGGAGCAAAATGGTTAAGGT 1651
QY 1427 GTATAAATCCATTAATTTCTTGGCCACATCTTCTGAGCCCTACTATTCAATTTGCTG 1486
Db 1652 GTATAAATCCATTAATTTCTTGGCCACATCTTCTGAGCCCTACTATTCAATTTGCTG 1711
QY 1487 TTTATATTCACAAAGCATGTGTCCAAAGAAAGAAAGTTAAGAGATGGAATATTCAT 1546
Db 1712 TTTATATTCACAAAGCATGTGTCCAAAGAAAGAAAGTTAAGAGATGGAATATTCAT 1771
QY 1547 TTCCTGTGTGGCTGTGTGGGGAATGTGTGCAAAACTGTCTCTTTCACAGCACT 1606
Db 1772 TTCCTGTGTGGCTGTGTGGGGAATGTGTGCAAAACTGTCTCTTTCACAGCACT 1831
QY 1607 CCTTTGCCCGAGAGAGAGATGGAAAGCCAGAGGAGTGGAGATGATGCTTCCAGCT 1666
Db 1832 CCTTTGCCCGAGAGAGATGGAAAGCCAGAGGAGTGGAGATGATGCTTCCAGCT 1891
QY 1667 GTGCTCTGTGTGCAAGCAAGTCTTATTTGGGCGCAAAAGGGGAAACTTTTGTGGAG 1726
Db 1892 GTGCTCTGTGTGCAAGCAAGTCTTATTTGGGCGCAAAAGGGGAAACTTTTGTGGAG 1950
QY 1727 AGGCGCTGTGCTTGTCAACCAAGTGGAAATGCAATGGGCGGATCTCAGCTCAGCCGAC 1786
Db 1951 AGGCGCTGTGCTTGTCAACCAAGTGGAAATGCAATGGGCGGATCTCAGCTCAGCCGAC 2010
QY 1787 CTCACCTCTGTGGGTCAAGTAATTTCTGCTCTGAGCCCTCCCAAGTAGCTGGGAATAGA 1846
Db 2011 CTCACCTCTGTGGGTCAAGTAATTTCTGCTCTGAGCCCTCCCAAGTAGCTGGGAATAGA 2070
QY 1847 GGCAGCCCAACATGCCCAAGTAATTTGTATTTTCAAGTAAGAAAGGGATTTCAACAGCT 1906
Db 2071 GGCAGCCCAACATGCCCAAGTAATTTGTATTTTCAAGTAAGAAAGGGATTTCAACAGCT 2130
QY 1907 TGGCCAGGCTGTGTAACCTGTAACCGCAAGTGAATCCACCGGCTCCGCTCCCAAG 1966
Db 2131 TGGCCAGGCTGTGTAACCTGTAACCGCAAGTGAATCCACCGGCTCCGCTCCCAAG 2190
QY 1967 TGTGGGATTAACAGGCTGAGCAGCCGCTGCGCCCAAGGGGAAACTTTGTGGAG 2026
Db 2191 TGTGGGATTAACAGGCTGAGCAGCCGCTGCGCCCAAGGGGAAACTTTGTGGAG 2250
QY 2027 AGCAGAGGGGCTCAATCTCCCTGTGATTTCCCAATGCAATGCTTATCTTCCCA 2086
Db 2251 AGCAGAGGGGCTCAATCTCCCTGTGATTTCCCAATGCAATGCTTATCTTCCCA 2310
QY 2087 TCTAGCAGGAATCTATTGTGTTTCTTCTGCAATTTAATAATGTAATGTGGCG 2146
Db 2311 TCTAGCAGGAATCTATTGTGTTTCTTCTGCAATTTAATAATGTAATGTGGCG 2370
QY 2147 CTACACACACCCCTGAGGAGGAGTGGAGAGGGTGGAGGCTGCTGCTGCTCACTT 2206
Db 2371 CTACACACACCCCTGAGGAGGAGTGGAGAGGGTGGAGGCTGCTGCTGCTCACTT 2430
QY 2207 TTTCTACCTTGAAGCTGATTAATAAATCACTTCTGTTGTTCAGTTTT 2258
Db 2431 TTTCTACCTTGAAGCTGATTAATAAATCACTTCTGTTGTTCAGTTTT 2482

RESULT 5
AX31926
LOCUS AX31926 1842 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 2435 from Patent WO0194629.
ACCESSION AX31926
VERSION AX31926.1 GI:18122560
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Hom sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS Young, P. E., Augustus, M., Carter, K. C., Ehner, R., Endress, G.,
Horrigan, S., Soppet, D. R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 2435 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source location/Qualifiers
1..1842
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 413 a 525 c 413 g 491 t
ORIGIN
Query Match 81.0%; Score 1842; DB 6; Length 1842;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 279 TACCTCATCACCCTCTTCATACCTTTACAGGCTCTCAATGCTATTATTTAATTTGA 338
Db 1 TACCTCATCACCCTCTTCATACCTTTACAGGCTCTCAATGCTATTATTTAATTTGA 60
QY 339 AACCACTTACACACTCCCTGCTGTGATTTGCTTCAAGTTCCATCCTTGCACTAATG 398
Db 61 AACCACTTACACACTCCCTGCTGTGATTTGCTTCAAGTTCCATCCTTGCACTAATG 120
QY 399 GGGCGACCATCACTGCGCTCTCACTACATCTTTGCTTCCAGATGGCTACCTTCTGCT 458
Db 121 GGGCGACCATCACTGCGCTCTCACTACATCTTTGCTTCCAGATGGCTACCTTCTGCT 180
QY 459 GGATCTATTACATGCGCAACCGCAATCAAGTATCAAGTGAATGCAATGCAATGTGTT 518
Db 181 GGATCTATTACATGCGCAACCGCAATCAAGTATCAAGTGAATGCAATGTGTT 240
QY 519 CTGACTTTGAACCTATTGCTTGGCTGTGACTACTTTGACGAGGAGAAATCAGAAAT 578
Db 241 CTGACTTTGAACCTATTGCTTGGCTGTGACTACTTTGACGAGGAGAAATCAGAAAT 300
QY 579 TCCCTGTCTCTGAGCAACAGAAATATGCAATAGTGTGTTCTTCCCTGCTGAAATT 638
Db 301 TCCCTGTCTCTGAGCAACAGAAATATGCAATAGTGTGTTCTTCCCTGCTGAAATT 360
QY 639 GCTGTTTCTCTACTCTTATAGGGGCTTCTTGTAAGGCCCCCAGTTCTCATATGATC 698
Db 361 GCTGTTTCTCTACTCTTATAGGGGCTTCTTGTAAGGCCCCCAGTTCTCATATGATC 420
QY 699 TACATGAAGCTGTGAGGAGAGCTGATTTGACATACAGAAAGATACCAACAGCATC 758
Db 421 TACATGAAGCTGTGAGGAGAGCTGATTTGACATACAGAAAGATACCAACAGCATC 480
QY 759 ATTCTGCTCTGAGCGCTGAGTGTGGCTTTTCTACTGATGAGGCTACACATGCTC 818
Db 481 ATTCTGCTCTGAGCGCTGAGTGTGGCTTTTCTACTGATGAGGCTACACATGCTC 540
QY 819 AACCCCAATACAGAAAGATATCTCTCATGAAAGTATGACAAACACCCCTTCTG 878
Db 541 AACCCCAATACAGAAAGATATCTCTCATGAAAGTATGACAAACACCCCTTCTG 600
QY 879 TTCCTGATGATGATGCTGATCTGAGGCAAGTTTGTGCTGTATCAATATGATGATCTGT 938

Db 601 TTCCGCTGACATGATGATCTGGGCAAGTTGTGCTGATCAATATATGACCTGT 660
Qy 939 TGGCTGGTCAAGAGAGATATGATTTTGAACGGGCTTGAATGGCTTTGAAGA 998
Db 661 TGGCTGGTCAAGAGAGATATGATTTTGAACGGGCTTGAATGGCTTTGAAGA 720
Qy 999 AAGGGCAAGGCAAGTGGGATGCTGTGCAACATGAAGTGTGCTTTGAACAAC 1058
Db 721 AAGGGCAAGGCAAGTGGGATGCTGTGCAACATGAAGTGTGCTTTGAACAAC 780
Qy 1059 CCCCCTTCACTGGACCATTTGCTCATTTCAATCAACCAAGGCTGGTGGCCGC 1118
Db 781 CCCCCTTCACTGGACCATTTGCTCATTTCAATCAACCAAGGCTGGTGGCCGC 840
Qy 1119 TACATCTTCAAGACACTCAAGTTCTTGAATTAAGAACTCTCTAGGGTCTCTG 1178
Db 841 TACATCTTCAAGACACTCAAGTTCTTGAATTAAGAACTCTCTAGGGTCTCTG 900
Qy 1179 CTATTCCTGGCCCTCTGGCAAGGCTGCACTCAGGATACCTGTCTGCTTCCAGATGAA 1238
Db 901 CTATTCCTGGCCCTCTGGCAAGGCTGCACTCAGGATACCTGTCTGCTTCCAGATGAA 960
Qy 1239 TTCCCTGTTGTTATTTGGAAGAGAGAGGCTGCAAGGCTCAATCAAGAGCCCAACCTG 1298
Db 961 TTCCCTGTTGTTATTTGGAAGAGAGAGGCTGCAAGGCTCAATCAAGAGCCCAACCTG 1020
Qy 1299 AGCAAGCTGGCCGCAATTAAGTCTCTCAGCCCTTCTACTATTGTTGCAACAGACCATC 1358
Db 1021 AGCAAGCTGGCCGCAATTAAGTCTCTCAGCCCTTCTACTATTGTTGCAACAGACCATC 1080
Qy 1359 CACTGGCTCTTCAATGGGTTACTCTCATATGCTGCTTCTGCTCTTCACTGGGCAATATG 1418
Db 1081 CACTGGCTCTTCAATGGGTTACTCTCATATGCTGCTTCTGCTCTTCACTGGGCAATATG 1140
Qy 1419 CTTAAGGTGATTAATTCATCTATTTCTGGGCAACATCTTCTGAGCTATCTATTC 1478
Db 1141 CTTAAGGTGATTAATTCATCTATTTCTGGGCAACATCTTCTGAGCTATCTATTC 1200
Qy 1479 ATATGCTTATATTCACAAAGCAATGATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1538
Db 1201 ATATGCTTATATTCACAAAGCAATGATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy 1539 TAATCCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1598
Db 1261 TAATCCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Qy 1599 ACAGCACTCTTGGCCCAAG 1658
Db 1321 ACAGCACTCTTGGCCCAAG 1380
Qy 1659 TTCCAGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1718
Db 1381 TTCCAGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
Qy 1719 TTTGGAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1778
Db 1441 TTTGGAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
Qy 1779 ACCGCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1838
Db 1501 ACCGCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
Qy 1839 GGAATACAGGCAAGCAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1898
Db 1561 GGAATACAGGCAAGCAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
Qy 1899 CACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1958
Db 1621 CACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
Qy 1959 TCCCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2018

Db 1681 TCCCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Qy 2019 GTGGAG 2078
Db 1741 GTGGAG 1800
Qy 2079 TCTCCCATCTAGCCAGAGATCTATGTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2120
Db 1801 TCTCCCATCTAGCCAGAGATCTATGTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1842

RESULT 6
AX410769 1842 bp DNA linear PAT 14-JUN-2002
LOCUS Sequence 3416 from Patent WO0229103.
DEFINITION AX410769
ACCESSION AX410769
VERSION AX410769.1 GI:21443474
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1
Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J. G.
Gene expression profiles in liver cancer
Patent: WO 0229103-A 3416 11-APR-2002;
GENE LOGIC INC (US)
Location/Qualifiers
FEATURES
source
1. 1842
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. U72515"
BASE COUNT 413 a 525 c 413 g 491 t
ORIGIN

Query Match 81.0%; Score 1842; DB 6; Length 1842;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 279 TACCTATCCACCTCTTCCATACCTTTACAGGCTCTCATTTCTATTTTAACTTTGA 338
Db 1 TACCTATCCACCTCTTCCATACCTTTACAGGCTCTCATTTCTATTTTAACTTTGA 60
Qy 339 AACCAAGCTTACCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 398
Db 61 AACCAAGCTTACCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
Qy 399 GGGCGACCATCACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 458
Db 121 GGGCGACCATCACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
Qy 459 GGATATCTTACACTGCGCAAGCACTAGATATCAAGTGAAGCAATGCAATGCTTGT 518
Db 181 GGATATCTTACACTGCGCAAGCACTAGATATCAAGTGAAGCAATGCTTGT 240
Qy 519 CTGACTTTGAAGCTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 578
Db 241 CTGACTTTGAAGCTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
Qy 579 TCCCTGCTCTGAGCAAGAAATATGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 638
Db 301 TCCCTGCTCTGAGCAAGAAATATGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
Qy 639 GCTGCTTCTCTACTTCTATGAGGCTCTTGTGTAAGGCTCCAGATTTCTCAATGAATCAC 698
Db 361 GCTGCTTCTCTACTTCTATGAGGCTCTTGTGTAAGGCTCCAGATTTCTCAATGAATCAC 420
Qy 699 TACATGAAGTGTGCAAGGAGAGCTGATTTGATATACAGAGAAAGATACCAACAGATC 758
Db 421 TACATGAAGTGTGCAAGGAGAGCTGATTTGATATACAGAGAAAGATACCAACAGATC 480
Qy 759 ATTCTGCTCTCAAGGCTGAGTCTGGGCTTTTCTACTAGTGGGCTACACATGCTC 818

Query Match 81.0%; Score 1842; DB 9; Length 1842;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

279 TACCTCATCAACCTCTTCCATTAACCTTACAGGCTCCATAGTGTATTTTAACCTTGA 338
1 TACTCATCAACCTCTTCCATTAACAGGCTCTCAATGCTTAATTTAACTTTGA 60

339 AACGAGCTTACACACTCCCTGCTGTATGTTGCTTCACTTCTCATCTTCAGTAATG 398
61 AACGAGCTTACACACTCCCTGCTGTATGTTGCTTCACTTCTCATCTTCAGTAATG 120

399 GGGCGACCATCACTGCGCTCTCACTAATCTTTGCTTCAAGATGCTTACCTTCTGCT 458
121 GGGCGACCATCACTGCGCTCTCACTAATCTTTGCTTCAAGATGCTTACCTTCTGCT 180

459 GGATACATTAACATGCGACCGGCACTAGATATCAAGTGGAGCAATGGCAATGTTGTT 518
181 GGATACATTAACATGCGACCGGCACTAGATATCAAGTGGAGCAATGGCAATGTTGTT 240

519 CTGACTTTGAAGCTGATGTTGGTGGCTGTGACTACTTGTGACGAGGGAAGATCAGAT 578
241 CTGACTTTGAAGCTGATGTTGGTGGCTGTGACTACTTGTGACGAGGGAAGATCAGAT 300

579 TCTCTGTCTCTGAGCAACAGAAATATGCCATACGTGTGTTCTTCTCTGCTGGAAGTT 638
301 TCTCTGTCTCTGAGCAACAGAAATATGCCATACGTGTGTTCTTCTCTGCTGGAAGTT 360

639 GCTGCTTTCT 698
361 GCTGCTTTCT 420

699 TACATGAAGCTGTGACGAGGAGAGCTGATGATACATACAGAGAAAGATACCAAGATC 758
421 TACATGAAGCTGTGACGAGGAGAGCTGATGATACATACAGAGAAAGATACCAAGATC 480

759 ATTCTCTCTCTCAAGCGCTGTGAGTCTGAGGCTTTTCTTCACTAATGCTGACACTGCT 818
481 ATTCTCTCTCTCAAGCGCTGTGAGTCTGAGGCTTTTCTTCACTAATGCTGACACTGCT 540

819 AGCCGCAATCAACAGAACTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 878
541 AGCCGCAATCAACAGAACTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600

879 TTCCGCTGATGATCACTGCTGATCTGAGGCAAGTTTGTCTGTACAAATATGTCACCTGT 938
601 TTCCGCTGATGATCACTGCTGATCTGAGGCAAGTTTGTGTCTGTACAAATATGTCACCTGT 660

939 TGGCTGTGACAGAAAGAGATGATGATTTGACGAGGCTGAGGCTTCAATGCTTGAAGAA 998
661 TGGCTGTGACAGAAAGAGATGATGATTTGACGAGGCTGAGGCTTCAATGCTTGAAGAA 720

999 AAGGGAAGGCAAGAGTGGAGTGGCTGTGCAACATGAAAGTGTGGCTTTTGAAGAAAC 1058
721 AAGGGAAGGCAAGAGTGGAGTGGCTGTGCAACATGAAAGTGTGGCTTTTGAAGAAAC 780

1059 CCCCCTTCACTGCGACCATTTGCTCTCAATCAACATCAACACGCTGAGTGGCGC 1118
781 CCCCCTTCACTGCGACCATTTGCTCTCAATCAACATCAACACGCTGAGTGGCGC 840

1119 TACATCTTCAAGACATCAAGTCTCTTGAAGAAATGAAGATCTCTCAGGCTCTCTGTT 1178
841 TACATCTTCAAGACATCAAGTCTCTTGAAGAAATGAAGATCTCTCAGGCTCTCTGTT 900

1179 CTATTTCTGCGCTCTGCGACGCTGCTGACATCAGGATACCTGCTGCTTCAAGTGA 1238
901 CTATTTCTGCGCTCTGCGACGCTGCTGACATCAGGATACCTGCTGCTTCAAGTGA 960

1239 TTCTCTCATTTGATTTGGAAGACAGAGCTCCAGGCTCATTTCAAGAGGCCACCTGT 1298
961 TTCTCTCATTTGATTTGGAAGACAGAGCTCCAGGCTCATTTCAAGAGGCCACCTGT 1020

QY 1299 AGCAAGCTGGCGCCCATTAAGTCTCTCAAGCCCTTCTATTTGGTGGCAACAGACATC 1358
DB 1021 AGCAAGCTGGCGCCCATTAAGTCTCTCAAGCCCTTCTATTTGGTGGCAACAGACATC 1080

QY 1359 CACTGCTCTTCAATGAGTACTCTCATGATGCTGCTTCTGCTCTTCAAGTGGCAATAG 1418
DB 1081 CACTGCTCTTCAATGAGTACTCTCATGATGCTGCTTCTGCTCTTCAAGTGGCAATAG 1140

QY 1419 CTTAAGGTATTAATCACTATTTCTTGGCCACATCTTCTCTGAGCTTCAATTC 1478
DB 1141 CTTAAGGTATTAATCACTATTTCTTGGCCACATCTTCTCTGAGCTTCAATTC 1200

QY 1479 ATATGCTCTTATTTTCAAGAAAGTGGCCCAAGAAAGAGAAAGTAAAGATGGA 1538
DB 1201 ATATGCTCTTATTTTCAAGAAAGTGGCCCAAGAAAGAGAAAGTAAAGATGGA 1260

QY 1539 TAATCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1598
DB 1261 TAATCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320

QY 1599 ACAGCACTCTTGGCCCAAGAGCAAGAAATGAAACCAAGAGGTGGAATGATGC 1658
DB 1321 ACAGCACTCTTGGCCCAAGAGCAAGAAATGAAACCAAGAGGTGGAATGATGC 1380

QY 1659 TTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1718
DB 1381 TTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440

QY 1719 TTGGAGAAAGCGCTTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1778
DB 1441 TTGGAGAAAGCGCTTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500

QY 1779 ACCGCAACTCTCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1838
DB 1501 ACCGCAACTCTCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560

QY 1839 GGAATACAGGCAAGCCACATGCGCCAGCTAATTTTGTATTTTCAAGAAAGGGAATT 1898
DB 1561 GGAATACAGGCAAGCCACATGCGCCAGCTAATTTTGTATTTTCAAGAAAGGGAATT 1620

QY 1899 CACCAGTGGGCGAGGCTGCTGCACTCTGCACTGCAAGTATCACCCTGCTGCTGCT 1958
DB 1621 CACCAGTGGGCGAGGCTGCTGCACTCTGCACTGCAAGTATCACCCTGCTGCTGCT 1680

QY 1959 TCCCAAGTCTGAGATTAACGCGGTGAGCACTGCTGCTGCTGCTGCTGCTGCTGCT 2018
DB 1681 TCCCAAGTCTGAGATTAACGCGGTGAGCACTGCTGCTGCTGCTGCTGCTGCTGCT 1740

QY 2019 GTGGAGAGAGCAAGGAGGCTCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2078
DB 1741 GTGGAGAGAGCAAGGAGGCTCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800

QY 2079 TCTCCCATCTAGCAGGAATCTATTTGTTTCTTCTGCT 2120
DB 1801 TCTCCCATCTAGCAGGAATCTATTTGTTTCTTCTGCT 1842

RESULT 8
AX073560 1461 bp DNA linear PAT 06-FEB-2001
Sequence 12 from Patent WO0104297.
AX073560.1 GI:12709973
AX073560.1
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Kato, S. and Kimura, T.
Human proteins having hydrophobic domains and dnas encoding these
proteins
JOURNAL Patent: WO 0104297-A 12 18-JAN-2001;

FEATURES
 SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Proteogene Inc. (JP)
 Location/Qualifiers
 1..1461
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 326 a 401 c 336 g 398 t
 ORIGIN

Query Match 64.3%; Score 1461; DB 6; Length 1461;
 Best Local Similarity 100.0%; Pred No. 0;
 Matches 1461; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 78 ATGGCGTCTCAGCGGAGGGGACGAGGGGACTGTGTGCGCGCTGCGGGGCTTCTGCAG 137
DB 1 ATGGCGTCTCAGCGGAGGGGACGAGGGGACTGTGTGCGCGCTGCGGGGCTTCTGCAG 60
QY 138 TGGGCTTTCAGAGAGCTGAGCTTAAACAAGTTGGCGAGCTCCCTGGGCGCTCAGAACAG 197
DB 61 TGGGCTTTCAGAGAGCTGAGCTTAAACAAGTTGGCGAGCTCCCTGGGCGCTCAGAACAG 120
QY 198 GGGCTGCGGCTGATCATCTCCATCTTCGGGGTTACCCCTTTGTTTATGGGAT 257
DB 121 GGGCTGCGGCTGATCATCTTCATCTTCGGGGTTACCCCTTTGTTTATGGGAT 180
QY 258 TACCTTTTACAGAGAGAGCTACCTCATACCTCTTCATACCTTTACAGGCTCTCA 317
DB 181 TACCTTTTACAGAGAGAGCTACCTCATACCTCTTCATACCTTTACAGGCTCTCA 240
QY 318 ATTGCTTATTTTAACTTTGGAAACAGCTCTTACCACTCCCTGCTGTGTAATTGCTTCA 377
DB 241 ATTGCTTATTTTAACTTTGGAAACAGCTCTTACCACTCCCTGCTGTGTAATTGCTTCA 300
QY 378 TTCTCATCTCTGACTAATAGGGCGGACATCATCTGCGCTCTCACTACCTTTGCTTC 437
DB 301 TTCTCATCTCTGACTAATAGGGCGGACATCATCTGCGCTCTCACTACCTTTGCTTC 360
QY 438 CAGATGCGCTTACCTTCTGCTGATATCTATTAACCTGACCGGACATACATCAAG 497
DB 361 CAGATGCGCTTACCTTCTGCTGATATCTATTAACCTGACCGGACATACATCAAG 420
QY 498 TGGACAATGCCACATTTGTTCTGACTTTGAAGCTGATTTGGCTGTGACTATT 557
DB 421 TGGACAATGCCACATTTGTTCTGACTTTGAAGCTGATTTGGCTGTGACTATT 480
QY 558 GACGGAGGGAAGATAGAAATTCCTGCTGAGCAACAGAAATATGCCATAGCTGT 617
DB 481 GACGGAGGGAAGATAGAAATTCCTGCTGAGCAACAGAAATATGCCATAGCTGT 540
QY 618 GTTCCCTCCCTGCTGGAAGTTGCTGCTTCTCTACTTCTATGGGCGCTTTCTTGTAAG 677
DB 541 GTTCCCTCCCTGCTGGAAGTTGCTGCTTCTCTACTTCTATGGGCGCTTTCTTGTAAG 600
QY 678 CCCCACTTCAATGATATCACTACATGAGCTGTGCGAGGAGAGCTGATGACATACCA 737
DB 601 CCCCACTTCAATGATATCACTACATGAGCTGTGCGAGGAGAGCTGATGACATACCA 660
QY 738 GGAAGATATCCAAACAGCATCTTCTGCTTCAAGCGCTGAGTGTGGGCTTTTCTAC 797
DB 661 GGAAGATATCCAAACAGCATCTTCTGCTTCAAGCGCTGAGTGTGGGCTTTTCTAC 720
QY 798 CTAGTGGCTTACACTGCTCAGCCGCCACATACAGAGACTATCTCTCACTGAAGAC 857
DB 721 CTAGTGGCTTACACTGCTCAGCCGCCACATACAGAGACTATCTCTCACTGAAGAC 780
QY 858 TATGACACACCCCTTCTGCTTCCGCTGCTGATGATGATCTGATGAGGCAAGTTGTG 917
DB 781 TATGACACACCCCTTCTGCTTCCGCTGCTGATGATGATCTGATGAGGCAAGTTGTG 840
QY 918 CTGTCAATATATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 977
DB 841 CTGTCAATATATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900

```

```

QY 978 GGCCTCAATGCTTTGAAAGAAAAGGCAAGGCAAGAGTGAGATGCTGTGCCAATGAAAG 1037
DB 901 GGCCTCAATGCTTTGAAAGAAAAGGCAAGGCAAGAGTGAGATGCTGTGCCAATGAAAG 960
QY 1038 GTGTGGCTTTTGAAGCAAAACCCCGCTTCACTGGACACCATGCTCATTTCAATCAAC 1097
DB 961 GTGTGGCTTTTGAAGCAAAACCCCGCTTCACTGGACACCATGCTCATTTCAATCAAC 1020
QY 1098 ACCAAGCGCTGGAGGCGCGCTACATCTTCAAGACATCAAGTCTCTTGGAAATAAGAA 1157
DB 1021 ACCAAGCGCTGGAGGCGCGCTACATCTTCAAGACATCAAGTCTCTTGGAAATAAGAA 1080
QY 1158 CTCTCTCAAGGCTCTGCTGTGCTATTCCTGAGCCCTCTGSCAGGCGCTGACCTCAGATAC 1217
DB 1081 CTCTCTCAAGGCTCTGCTGTGCTATTCCTGAGCCCTCTGSCAGGCGCTGACCTCAGATAC 1140
QY 1218 CTGTCTGCTTTCAGATGGAATTCCTCATTTGTTATTTGGAAGAACAGGCTGCCAGGCTC 1277
DB 1141 CTGTCTGCTTTCAGATGGAATTCCTCATTTGTTATTTGGAAGAACAGGCTGCCAGGCTC 1200
QY 1278 ATTCAAGAGAGCCCAACCTGAGCAAGCTGGCGGCATTACCTCCAGGCCCTTTCAC 1337
DB 1201 ATTCAAGAGAGCCCAACCTGAGCAAGCTGGCGGCATTACCTCCAGGCCCTTTCAC 1260
QY 1338 TATTTGGTCAACAGACCATCCACTGCTCTTCAATGGGTTACTCCATGACTGCTTCTGAC 1397
DB 1261 TATTTGGTCAACAGACCATCCACTGCTGCTCTTCAATGGGTTACTCCATGACTGCTTCTGAC 1320
QY 1398 CTCTTCAAGTGGGCAAAATGCTTAAAGTGTATTAATTCATTTATTTCTTGGCCATAC 1457
DB 1321 CTCTTCAAGTGGGCAAAATGCTTAAAGTGTATTAATTCATTTATTTCTTGGCCATAC 1380
QY 1458 TTCTTCTGAGCTCTATTCATATTTGCTTATTTCAACAAAGCAATGTCGCAAGGAA 1517
DB 1381 TTCTTCTGAGCTCTATTCATATTTGCTTATTTCAACAAAGCAATGTCGCAAGGAA 1440
QY 1518 GAGAAGTTAAAGAGATGAA 1538
DB 1441 GAGAAGTTAAAGAGATGAA 1461

```

RESULT 9
 AK058063 1573 bp mRNA linear PRI 31-OCT-2001
 LOCUS
 DEFINITION
 AK058063
 ACCESSION
 AK058063.1 GI:16554077
 VERSION
 oligo capping; fls (full insert sequence).
 KEYWORDS
 Homo sapiens (human)
 SOURCE
 Homo sapiens
 ORGANISM

REFERENCE
 1
 Ishibashi, T., Kanehori, K., Yosida, M., Matanabe, S., Ishida, S.,
 Ono, Y., Hociuta, T., Hiraoka, S., Murakawa, K., Takiguchi, S.,
 Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M.,
 Yamashita, H., Chiba, Y., Suzuki, Y., Hata, H., Nakagawa, K., Mizuno, S.,
 Morinaga, M., Kawamura, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,
 Nishikawa, T., Sugiyama, A., Kawakami, B., Nagai, K., Isogai, T. and
 Sugano, S.
 NEDO human cDNA sequencing project

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing;
 Research Association for Biotechnology (RAB); cDNA library

constructi and 5'-end one pass sequencing: institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; 3'-end one pass sequencing: RAB, clone selection for full insert sequencing: RAB and Helix Research Institute.

FEATURES	Location/Qualifiers
SOURCE	1. .1573

BASE COUNT	358 a	436 c	378 g	401 t
ORIGIN				

Query March	54.5%;	Score 1238;	DB 9;	Length 1573;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 1388;	Conservative	0;	Mismatches 3;	Indels 0;
				Gaps 0;

QY	864	AAACACCCCTTCTGGTTCGCGTGAATGAATGCTGAATCTGGGGCAATTGTCTCTAC	923
Dd	183	AAACACCCCTTCTGGTTCGCGTGAATGAATGCTGAATCTGGGGCAATTGTCTCTAC	242
QY	924	AAATATGTACCTGTGGCTGGTCAAGAGAAGTATGACTATTTGACGGGCTTGGGCTTC	983
Dd	243	AAATATGTACCTGTGGCTGGTCAAGAGAAGTATGACTATTTGACGGGCTTGGGCTTC	302
QY	984	AATGGCTTGAAGAAAGGGCAAGGCCAAGTGGATGCTGTGGCAACATGAAGTGTGG	1043
Dd	303	AATGGCTTGAAGAAAGGGCAAGGCCAAGTGGATGCTGTGGCAACATGAAGTGTGG	362
QY	1044	CTCTTTGAAACAAACCCCGGCTTCACTGGACCAATGGCTCATTCACATCAACCAAC	1103
Dd	363	CTCTTTGAAACAAACCCCGGCTTCACTGGACCAATGGCTCATTCACATCAACCAAC	422
QY	1104	GCCTTGGTGGCCCGCTACATCTTCAACAGACTCAAATTCTTTGGAAATTAAGAACTCTCT	1163
Dd	423	GCCTTGGTGGCCCGCTACATCTTCAACAGACTCAAATTCTTTGGAAATTAAGAACTCTCT	482
QY	1164	CAGGCTCTCTGTTGCTATTCCTGGCCCTGTGGCACGGCGCTGCACCTCAGATPACTGGTC	1223
Dd	483	CAGGCTCTCTGTTGCTATTCCTGGCCCTGTGGCACGGCGCTGCACCTCAGATPACTGGTC	542
QY	1224	TGCTTCCAGATGGAATTCCTCATTTGTATTTGTGGAAACACAGGCTGCCAGGCTCATTCAA	1283
Dd	543	TGCTTCCAGATGGAATTTCTCATTTGTATTTGTGGAAACACAGGCTGCCAGGCTCATTCAA	602
QY	1284	GAGAGCCCCACCTGAGCAAGCTGGCGGCATTACTGTCTTCACAGCCCTTCTACTATTTG	1343
Dd	603	GAGAGCCCCACCTGAGCAAGCTGGCGGCATTACTGTCTTCACAGCCCTTCTACTATTTG	662
QY	1344	GTGCACACAGACCATCATCGGTCTTCAATGGGTTACTGCATGACAGCTTCGTGGCTCTTC	1403
Dd	663	GTGCACACAGACCATCATCGGTCTTCAATGGGTTACTGCATGACAGCTTCGTGGCTCTTC	722
QY	1404	ACGTGGGCAAAATGGCTTAAGGTGTATAAATCATCTATTTTCTTGGCACATCTTCTTC	1463
Dd	723	ACGTGGGCAAAATGGCTTAAGGTGTATAAATCATCTATTTTCTTGGCACATCTTCTTC	782
QY	1464	CTGAGCCTACTATTCATATTTGCTTATATTCACAAGCAAGTGTCCAGAAGAAAGAGAG	1523
Dd	783	CTGAGCCTACTATTCATATTTGCTTATATTCACAAGCAAGTGTCCAGAAGAAAGAGAG	842
QY	1524	TTAAAGAAAGATGGAATTAATCCATTTTCCCTGGTGGCTGTGGGGGCACTGTGCAGAACTA	1583
Dd	843	TTAAAGAAAGATGGAATTAATCCATTTTCCCTGGTGGCTGTGGGGGCACTGTGCAGAACTA	902
QY	1584	CTCGTCTCCCTTTTACACAGCACTCCTTTGGCCCCACAGACAGAAATGAGAAAAGCCAGGAG	1643
Dd	903	CTCGTCTCCCTTTTACACAGCACTCCTTTGGCCCCACAGACAGAAATGAGAAAAGCCAGGAG	962
QY	1644	GTGGAAGATCGATGCTTCAGCTGTGGCTCTGTCTGCCAGCAAGTCTTCATTTTGGGGCCA	1703

Db	963	GTGGAAATCGATGCTTCCAGCTGTGTGCTCTGCTGCACGACCAAGTCTTCATTGGGGSCCA	1022
Oy	1704	AAGGGGAACTTTTTTTTGGAGAAAGGCGCTTGTGTGTACCCACGCTGGAAATGCAGTG	1763
Db	1023	AAGGGGAACTTTTTTTTGGAGAAAGGCGCTTGTGTGTACCCACGCTGGAAATGCAGTG	1082
Oy	1764	GCGGGATCTCAGCTCAGCCGCAACCTCCACCTCTGGGTTTCAGTGATTTTCTGCTCAG	1823
Db	1083	GCGGGATCTCAGCTCAGCCGCAACCTCCACCTCTGGGTTTCAGTGATTTTCTGCTCAG	1142
Oy	1824	CTTCCCAAGTACGTGGGAATACAGGCAACGCCACATGCCAGCTAATTTTTGTATTTTCA	1883
Db	1143	CTTCCCAAGTACGTGGGAATACAGGCAACGCCACATGCCAGCTAATTTTTGTATTTTCA	1202
Oy	1884	GTAGAAACGGGATTTTCCACAGTTGGGCGAGGCTGTGTCCAACTCTGACCGCAAGTAT	1943
Db	1203	GTAGAAACGGGATTTTCCACAGTTGGGCGAGGCTGTGTCCAACTCTGACCGCAAGTAT	1262
Oy	1944	CCACCCGCTCTCCGCTTCCCAAAGTGTGGGATTAACAGGCGTGAECACCGTGCCTGGGCC	2003
Db	1263	CCACCCGCTCTCCGCTTCCCAAAGTGTGGGATTAACAGGCGTGAECACCGTGCCTGGGCC	1322
Oy	2004	AAAGGGGAACTCTGTGGAGGAGGAGGAGGAGGCGTCACTCTCCCTGTGATTTCCCCAT	2063
Db	1323	AAAGGGGAACTCTGTGGAGGAGGAGGAGGAGGCGTCACTCTCCCTGTGATTTCCCCAT	1382
Oy	2064	GCACATGCTCTTATCTCTCCCATCTAGCCAGGAATCTATGTGTCTTCTTCTGCCAAT	2123
Db	1383	GCACATGCTCTTATCTCTCCCATCTAGCCAGGAATCTATGTGTCTTCTTCTGCCAAT	1442
Oy	2124	TTACTATGATTTGATGTGTGTGCGGTACACACACCCCCCATGGGGGGGTGGAGAGGGGT	2183
Db	1443	TTACTATGATTTGATGTGTGTGCGGTACACACACCCCCCATGGGGGGGTGGAGAGGGGT	1502
Oy	2184	GCAAGGCGCTGCTGCTCACAATTTTCTACCTTGGAACTGTATTAGTAAATCATTCT	2243
Db	1503	GCAAGGCGCTGCTGCTCACAATTTTCTACCTTGGAACTGTATTAGTAAATCATTCT	1562
Oy	2244	GTTTGTTCAGT 2254	
Db	1563	GTTTGTTCAGT 1573	

[illegible]

forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'Acc' after SalI site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame. Clone distribution: <http://bioinfo.clontech.com/orfclones>.

FEATURES

source

```

1. 1146
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GH00356X1.0"
/clone_lib="BD Creator(TM) CDS Library derived from MGC collection"
/lab_host="DH5alpha TI resistant"
/note="Vector: pDNR-Dual"
1. 1146
/codon_start=1
/product="putative protein similar to nesey (Drosophila)"
/protein_id="AAP35646.1"
/db_xref="GI:30582839"
/translation="MGRITVAVLTTCFQMAVILAGYVYTATGNDIKTMPCVLTLLKILGLAVDFDGDQNSLSBQKXAIKGLVPSLEAVSFYFGAFVLPQSMNHYMKLVQELIDIPKIPNSIIPALKRLSLGLFYLVGYTLSPHITBEDYLDNHPFERCMYMLIMGKFLVKYVLCWLVTEGVCILGNGEENGKAKMDCAKMKWLFENPRMTGILASFININAMVARIYFKRLKFLNGKLSGLIFLALHGHGSLVCFQMKLIIVVERQARLIQESPLSKLAIVLOPFYLVNOQTIHMLMGSMTRFCLTWDKMLVYSIYIFLGHIFPLSLFPIFYHKAWPKKKLKKME"

```

CDS

BASE COUNT 274 a 310 c 256 g 306 t

ORIGIN

Query Match Best Local Similarity 48.1%; Score 1094; DB 9; Length 1146; Mismatches 1144; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 396 ATGGGCGGACCATCACTGCGCTCTCACTACCTTTTGTCCAGATGGCTACCTTTCTG
DB 1 ATGGGCGGACCATCACTGCGCTCTCACTACCTTTTGTCCAGATGGCTACCTTTCTG 60
QY 456 GCTGATTAATTACACTGCGACCGGCACTAGATATCAAGTGAACATGCCAATTGT 515
DB 61 GCTGATTAATTACACTGCGACCGGCACTAGATATCAAGTGAACATGCCAATTGT 120
QY 516 GTTCTGACTTTGAAGTGAATGTTGGTGTGACTTCTGACGAGGAGGAAGATCAG 575
DB 121 GTTCTGACTTTGAAGTGAATGTTGGTGTGACTTCTGACGAGGAGGAAGATCAG 180
QY 576 AATTCCTGTCTCTAGCAACGAATAATGACATAGTGTGTTCTTCCCTGCTGAA 635
DB 181 AATTCCTGTCTCTAGCAACGAATAATGACATAGTGTGTTCTTCCCTGCTGAA 240
QY 636 GTTGTGCTTTCTCTCACTTCTATGGGGCTTCTTGTGTAAGGGCCCAAGTTTCATGAAT 695
DB 241 GTTGTGCTTTCTCTCACTTCTATGGGGCTTCTTGTGTAAGGGCCCAAGTTTCATGAAT 300
QY 696 CACTACATGAAGTGTGAGGAGGAGCTGATGATACACAGGAAGAATACCAACACAG 755
DB 301 CACTACATGAAGTGTGAGGAGGAGCTGATGATACACAGGAAGAATACCAACACAG 360
QY 756 ATCATTTCTGCTCTCAAGCGCTAGTGTGGGCTTTTCTACCTAGTGGGCTACACATG 815
DB 361 ATCATTTCTGCTCTCAAGCGCTAGTGTGGGCTTTTCTACCTAGTGGGCTACACATG 420
QY 816 CTCAGGCCCCCATCAACAGAAAGCTATCTCTCACTGAAGACTATGACCAACCCCTTC 875
DB 421 CTCAGGCCCCCATCAACAGAAAGCTATCTCTCACTGAAGACTATGACCAACCCCTTC 480
QY 876 TGGTTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 935
DB 481 TGGTTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
QY 936 TGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 995

```

```

DB 541 TGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
QY 996 GAAAAGGGGAGGCAAGAGTGGATGCTGTGCAACATGAAAGTGTGCTTTGAAACA 1055
DB 601 GAAAAGGGGAGGCAAGAGTGGATGCTGTGCAACATGAAAGTGTGCTTTGAAACA 660
QY 1056 AACCCCGGCTTACAGGACCACTTGGCTCATTCATCAATCAACCAAGCCCTGGGAGCC 1115
DB 661 AACCCCGGCTTACAGGACCACTTGGCTCATTCATCAATCAACCAAGCCCTGGGAGCC 720
QY 1116 CGCTACATTTTCAAAAGCACTCAAGTTCTTGGAAATTAAGAACTCTTCAGGGCTCTCG 1175
DB 721 CGCTACATTTTCAAAAGCACTCAAGTTCTTGGAAATTAAGAACTCTTCAGGGCTCTCG 780
QY 1176 TTGCTATTTCTGCGCTCTGCGACAGGCTGCACTCAGATATCTGTCTCTTCAGATG 1235
DB 781 TTGCTATTTCTGCGCTCTGCGACAGGCTGCACTCAGATATCTGTCTCTTCAGATG 840
QY 1236 GAATTCCTCATTTGTTATTTGTAAGAGACAGGCTGCGACGCTCATTCAGAGAGCCACCC 1295
DB 841 GAATTCCTCATTTGTTATTTGTAAGAGACAGGCTGCGACGCTCATTCAGAGAGCCACCC 900
QY 1296 CTGAGCAAGCTGCGCGCATTACTGCTCTCAAGCCCTTCTACTATTTGTGCAACAGACC 1355
DB 901 CTGAGCAAGCTGCGCGCATTACTGCTCTCAAGCCCTTCTACTATTTGTGCAACAGACC 960
QY 1356 ATCCACTGCTCTTCAATGGGTTACTGATGATGCTTGTGCTTCTTCACTGCGACAA 1415
DB 961 ATCCACTGCTCTTCAATGGGTTACTGATGATGCTTGTGCTTCTTCACTGCGACAA 1020
QY 1416 TGGCTTAAGTGTATTAATCCATCTATTCTCTGGCAGACATCTTCTCTGAGCCACTA 1475
DB 1021 TGGCTTAAGTGTATTAATCCATCTATTCTCTGGCAGACATCTTCTCTGAGCCACTA 1080
QY 1476 TTCATATTTGCCCTTAATATTCACAAAGCAATGATGCCAAGGAAAGAGTTAAAGATG 1535
DB 1081 TTCATATTTGCCCTTAATATTCACAAAGCAATGATGCCAAGGAAAGAGTTAAAGATG 1140
QY 1536 GAATA 1540
DB 1141 GAATA 1145

```

RESULT 11
BT007735
LOCUS 1146 bp mRNA linear SYN 13-MAY-2003
DEFINITION Synthetic construct Homo sapiens putative protein similar to nesey (Drosophila) mRNA, partial cds.
ACCESSION BT007735
VERSION BT007735.1 GI:30584308
KEYWORDS FLI-CDNA.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 1146)
AUTHORS Kalline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,D., Lin,Y., Pheasant,M., and Farmer,A.
TITLE Cloning of human full-length CDSs in BD Creator(TM) system donor vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1146)
AUTHORS Kalline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,D., Lin,Y., Pheasant,M., and Farmer,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal

tag). The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame. clone distribution: <http://bioinfo.clontech.com/orfclones>.

FEATURES

source

1. 1146
 /organism="synthetic construct"
 /mol_type="mRNA"
 /db_xref="taxon:32630"
 /clone="GH003561.0"
 /clone_1lb="BD Creator(TM) CDS Library derived from MGC collection"
 /lab_host="DH5alpha T1 resistant"
 /note="Vector: pDNR-Dual"
 1. >1146
 /note="Mutations: 1145:Stop->Leu"
 /codon_start=1
 /transl_table=1
 /product="Homo sapiens putative protein similar to nesey (Drosophila)"
 /protein_id="AAP36403.1"
 /db_xref="GI:30584309"
 /translation="MGRTTAVLTTPGFMAYLLAGVYATATGNYDKMTMPHCVLT.
 KLGLAVDFDQKDNLSSEQKXIRQVPSLLEVAGSYRFGATLVGPQSMNH
 MKLVQGLIDIPKIPNSIIIPALRSLSLGFLVAGYLLSPHLEDDVNH
 WFRGMVLMIGKFLYKVTWCMTLWTEGVCLTGLGFGFEEKAKADCAAMKVMLE
 ETNPRFGTSLFNINAWAVYIRKLFELSLQSLSLFLMLHGLHSGYLV
 CFPMKFLIVEROARLIOESPLSLKLAITVLOPYVLOOTIHLFPGYSMTATC
 LFTMDKMKLYKYSIYPLGHIFPLSLFPLIYHKAMVPRKXKMKML"

CDS

BASE COUNT 273 a 310 c 256 g 307 t
 ORIGIN

Query Match 48.1%; Score 1093; DB 12; Length 1146;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 396 ATGGGCGGACCATCATCTGCGGCTCTCACTACCTTTTGTTCAGATGGCTTACTTGTG 455
 Db 1 ATGGGCGGACCATCATCTGCGGCTCTCACTACCTTTTGTTCAGATGGCTTACTTGTG 60
 QY 456 GCTGGATCATATTACCTGCCACCGGCACTAGCATATCAAGTGGAGCAATGCCATTGT 515
 Db 61 GCTGGATCATATTACCTGCCACCGGCACTAGCATATCAAGTGGAGCAATGCCATTGT 120
 QY 516 GTTCTGACTTGAAGTGAATGTTGGCTGTGACTACTTGTGAAGAGGAAAGATCAG 575
 Db 121 GTTCTGACTTGAAGTGAATGTTGGCTGTGACTACTTGTGAAGAGGAAAGATCAG 180
 QY 576 AATTCCTTGTCTCTGAGCAACAGAAATATGCCATAGTGTGTTCTCTCTCTGGAA 635
 Db 181 AATTCCTTGTCTCTGAGCAACAGAAATATGCCATAGTGTGTTCTCTCTCTGGAA 240
 QY 636 GTTCTGCTTCTCTCTCTCTCTATGGGCGCTTGTGGAGGCGCCAGTTCTCATGAT 695
 Db 241 GTTCTGCTTCTCTCTCTCTCTATGGGCGCTTGTGGAGGCGCCAGTTCTCATGAT 300
 QY 696 CACTACATGAAGCTGTGACGAGAGCTGATTGACATACAGAGAAAGATACCAACAGC 755
 Db 301 CACTACATGAAGCTGTGACGAGAGAGCTGATTGACATACAGAGAAAGATACCAACAGC 360
 QY 756 ATCATCTCTGCTCAAGCGCTGAGTCTGGGCTTTTCTAAGTGGGCTTACACTG 815
 Db 361 ATCATCTCTGCTCAAGCGCTGAGTCTGGGCTTTTCTAAGTGGGCTTACACTG 420
 QY 816 CTCAGCCCCCAGACAGAGAGACTATCTCTCACTGAAGACTAGCAACACCCCTTC 875
 Db 421 CTCAGCCCCCAGACAGAGAGACTATCTCTCACTGAAGACTAGCAACACCCCTTC 480
 QY 876 TGGTTCGCTGCATGATCATGCTGATGCGGCAAGTTTGTCTGTACAATATGTCAAC 935
 Db 481 TGGTTCGCTGCATGATCATGCTGATGCGGCAAGTTTGTCTGTACAATATGTCAAC 540

QY 936 TGTTCCTGTGTCAAGAGAGATGATCATTTTGAACGGGCTTGAGCTTCAATGCTTTGAA 995
 Db 541 TGTTCCTGTGTCAAGAGAGATGATCATTTTGAACGGGCTTGAGCTTCAATGCTTTGAA 600
 QY 996 GAAAAGGCAAGGCAAGAGTGGAGTGGCTTGCCCATGATGAAGGTGGCTTTGAACA 1055
 Db 601 GAAAAGGCAAGGCAAGAGTGGAGTGGCTTGCCCATGATGAAGGTGGCTTTGAACA 660
 QY 1056 AACCCCGCTTCACTGGCAACCATTTGCTTATTCATCAATCAACCAACGCTGGTGCC 1115
 Db 661 AACCCCGCTTCACTGGCAACCATTTGCTTATTCATCAATCAACCAACGCTGGTGCC 720
 QY 1116 CGCTACATCTTCAAGACATGACATGCTTGTGAAATTAAGAAATCTCTCAAGGCTCTCG 1175
 Db 721 CGCTACATCTTCAAGACATGACATGCTTGTGAAATTAAGAAATCTCTCAAGGCTCTCG 780
 QY 1176 TTGCTATTCCTGACCTCTGGACAGGCTTGACATCTAGATATCTGCTTCCAGATG 1235
 Db 781 TTGCTATTCCTGACCTCTGGACAGGCTTGACATCTAGATATCTGCTTCCAGATG 840
 QY 1236 GAATTCCTCATTTGTTATTTGGAAGACAGGCTGCCAGGCTCATTAAGAGAGCCACAC 1295
 Db 841 GAATTCCTCATTTGTTATTTGGAAGACAGGCTGCCAGGCTCATTAAGAGAGCCACAC 900
 QY 1296 CTGAGCAAGTGGCGCCCATTAATGCTTCTTCAACCTTCTACTATTGTGCAACAGAC 1355
 Db 901 CTGAGCAAGTGGCGCCCATTAATGCTTCTTCAACCTTCTACTATTGTGCAACAGAC 960
 QY 1356 ATCCACTGCTCTTCAATGAGTTACTCATGATGCTTCTGCTTCAAGTGGGCAAA 1415
 Db 961 ATCCACTGCTCTTCAATGAGTTACTCATGATGCTTCTGCTTCAAGTGGGCAAA 1020
 QY 1416 TGGCTTAAGGTATTAATTCATCTATTTCTTGGCCACATCTTCTCTGACCTTACTA 1475
 Db 1021 TGGCTTAAGGTATTAATTCATCTATTTCTTGGCCACATCTTCTCTGACCTTACTA 1080
 QY 1476 TTGATATGCTTATTAATTCACAAAGCAATGCTGCCAAGAAAGAAATTAAGAGATG 1535
 Db 1081 TTGATATGCTTATTAATTCACAAAGCAATGCTGCCAAGAAAGAAATTAAGAGATG 1140
 QY 1536 GAAT 1539
 Db 1141 GAAT 1144

RESULT 12
 AC006512/c 155975 bp DNA linear PRI 04-APR-2003
 LOCUS Homo sapiens 12 PAC RP3-461P17 (Roswell Park Cancer Institute Human
 DEFINITION PAC library) complete sequence.
 ACCESSION AC006512
 VERSION AC006512.13 GI:29469488
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 1 (bases 1 to 155975)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Mammalia; Euthelial; Primates; Catarrhini; Homnidae; Homo.
 Alshrooke,S.L., Amaratunge,H.C., Aze,J.R., Ayele,M., Banks,T.,
 Barbara,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D.,
 Bouch,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhey,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dabholne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Hochs,S., Durbin,K.J.,
 Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emelting,S.,
 Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,
 Franz,P., Gabisi,A., Geo,J., Garcia,A., Garner,T., Garza,N.,

Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P., Hernandez, J., Hernandez, O., Hodgson, A., Hoggus, M., Holloway, C., Hollins, B., Homel, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L. E., Jacobson, B., Jia, T., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C., Kratovic, J., Kuresti, A., Landry, N., Leal, B., Lee, E., Lewis, L. C., Lewis, L., Li, J., Li, Z., Licharge, O., Liu, C., Liu, J., Liu, W., Louised, H., Losado, R. X., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Maronde, I., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M. P., Meador, M., Mel, G., Metcalf, T., Metzger, K., Montgomery, K. T., Morgan, M., Morris, S., Moser, M., Neahab, K., Neaton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokweto, S., Ogun, M., Okunnu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojubo, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shim, C., Shoohtari, N., Sison, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansy, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umanu, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Kuchelapati, R., Weinstein, G., and Gibbs, R.

TITLE
Unpublished

JOURNAL
Unpublished

REFERENCE
2 (bases 1 to 155975)

AUTHORS
Worley, K.C.

TITLE
Submitted (05-FEB-1999) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

JOURNAL
Submitted (05-FEB-1999) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
3 (bases 1 to 155975)

AUTHORS
Worley, K.C.

TITLE
Submitted (29-MAY-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

JOURNAL
Submitted (29-MAY-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
4 (bases 1 to 155975)

AUTHORS
Worley, K.C.

TITLE
Submitted (02-JUN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

JOURNAL
Submitted (02-JUN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
5 (bases 1 to 155975)

AUTHORS
Worley, K.C.

TITLE
Submitted (03-JUN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

JOURNAL
Submitted (03-JUN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
6 (bases 1 to 155975)

AUTHORS
Worley, K.C.

TITLE
Submitted (18-JUN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

JOURNAL
Submitted (18-JUN-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
7 (bases 1 to 155975)

AUTHORS
Worley, K.C.

TITLE
Submitted (27-OCT-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

JOURNAL
Submitted (27-OCT-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
8 (bases 1 to 155975)

AUTHORS
Worley, K.C.

TITLE
Submitted (29-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

JOURNAL
Submitted (29-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
9 (bases 1 to 155975)

AUTHORS
Worley, K.C.

TITLE
Direct Submission

JOURNAL
Submitted (02-APR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
10 (bases 1 to 155975)

AUTHORS
Worley, K.C.

TITLE
Submitted (04-APR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

JOURNAL
Submitted (04-APR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT
On Apr 2, 2003 this sequence version replaced gi:4926863.
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://www.hgsc.bcm.tmc.edu:8088/quality_info/genbank_annotation.html.

FEATURES

source
1. 155975
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP3-461F17"
1. 97210
/note="overlaps bases 127585..222930 of clone"
/function="clone overlap"

misc_feature
1. 97210
/note="overlaps bases 127585..222930 of clone"

repeat_region
1649..1772
/rpt_family="L3b"
complement(2781..3065)
/rpt_family="AluYd2"

repeat_region
3127..3430
/rpt_family="AluX"

STS
3473..4048
/standard_name="GDB:214837"

repeat_region
3592..3797
/rpt_family="MIR"

repeat_region
4550..4771
/rpt_family="L3b"

repeat_region
complement(4816..5127)
/rpt_family="AluSd"

repeat_region
complement(5146..5288)

```

repeat_region      /rpt_family="FLAM.C"
complement(5322..5506)
/rpt_family="AlusX"
repeat_region      5541..5833
/rpt_family="AluYb"
STS                6770..6912
/standard_name="GDB:384728"
STS                7448..7601
/standard_name="D12S1899"
STS                7675..7889
/standard_name="G10506"
repeat_region      7717..7746
/rpt_family="TTTA.n"
STS                7769..7860
/standard_name="SHG-58828"
repeat_region      9529..9570
/rpt_family="(TG).n"
complement(10903..11196)
/rpt_family="AlusX"
repeat_region      12008..12158
/rpt_family="CT-rich"
repeat_region      12166..12206
/rpt_family="(TCTCCC).n"
repeat_region      12368..12585
/rpt_family="(CGGG).n"
repeat_region      12647..12685
/rpt_family="GC-rich"
repeat_region      12822..12850
/rpt_family="GC-rich"

```

Query Match 31.1%; Score 708; DB 9; Length 155975;

Best Local Similarity 100.0%; Pred. No. 0; Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1553 GGTGGCTGTGCGGAGCTGTGCGAAGACTACTGCTCTCCCTTTTCACAGCACTCTTTG 1612
DB 61322 GGTGGCTGTGCGGAGCTGTGCGAAGACTACTGCTCTCCCTTTTCACAGCACTCTTTG 61263
QY 1613 CCCCAGAGCAGAGATGGAAGAGCAGGAGGTGGAAGATGATGCTTCCAGCTGTGCTT 1672
DB 61262 CCCCAGAGCAGAGATGGAAGAGCAGGAGGTGGAAGATGATGCTTCCAGCTGTGCTT 61203
QY 1673 CTGTGCGAGCAGAGCTTCTTATTTTGGGGCCAAAGGGGAACTTTTGGAGAAGGCT 1732
DB 61202 CTGTGCGAGCAGAGCTTCTTATTTTGGGGCCAAAGGGGAACTTTTGGAGAAGGCT 61143
QY 1733 CTGTGCTGTGACCCAGCTGGAATGCAAGTGGCGGATCTCAGCTCAACCGCAACTCTCAC 1792
DB 61142 CTGTGCTGTGACCCAGCTGGAATGCAAGTGGCGGATCTCAGCTCAACCGCAACTCTCAC 61083
QY 1793 CTCTGGGTTCAAGTATTTCTGCTCAGCTTCCCAAGTACGTTGGGAATACAGCAAG 1852
DB 61082 CTCTGGGTTCAAGTATTTCTGCTCAGCTTCCCAAGTACGTTGGGAATACAGCAAG 61023
QY 1853 CCACCATGCCAGGTAATTTTGTATTTTGTAGTAAGAAAGGAGTTTACCAAGCTTGCGCA 1912
DB 61022 CCACCATGCCAGGTAATTTTGTATTTTGTAGTAAGAAAGGAGTTTACCAAGCTTGCGCA 60963
QY 1913 GGCTGTGCTGCAATCTCTGACCGCAAGTATCAACCGCTTCCGCTCCCAAGTGTGCG 1972
DB 60962 GGCTGTGCTGCAATCTCTGACCGCAAGTATCAACCGCTTCCGCTCCCAAGTGTGCG 60903
QY 1973 GATTACAGGGCTGAGCCACCGTGCCTGGCCCAAGGGGAACTCTTGTGGAGAGACAGA 2032
DB 60902 GATTACAGGGCTGAGCCACCGTGCCTGGCCCAAGGGGAACTCTTGTGGAGAGACAGA 60843
QY 2033 GGGGCTACATCTCCCTCTGATCCCGCATGACATTCCTTATCTCTCCCATCTAGC 2092
DB 60842 GGGGCTACATCTCCCTCTGATCCCGCATGACATTCCTTATCTCTCCCATCTAGC 60783
QY 2093 CAGGAATCTATGTTGTTTCTCTGCAATTTACTATGATGTAATGTCGCCCTACCA 2152
DB 60782 CAGGAATCTATGTTGTTTCTCTGCAATTTACTATGATGTAATGTCGCCCTACCA 60723

```

```

QY 2153 CACACCCCCCATGAGGGGGGTGAGAGGGGTGCAAGGCCCTGCTCCACTTTTCTA 2212
DB 60722 CACACCCCCCATGAGGGGGGTGAGAGGGGTGCAAGGCCCTGCTCCACTTTTCTA 60663
QY 2213 CCTTGAAGCTGATATGATTAATCACTCTGTTGTTGCAAGTTTCA 2260
DB 60662 CCTTGAAGCTGATATGATTAATCACTCTGTTGTTGCAAGTTTCA 60615

```

RESULT 13
HSU47924/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE

1 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

2 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

3 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

4 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

5 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

6 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

7 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

8 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

9 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

10 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

11 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

12 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

13 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

14 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

15 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

16 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

17 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

18 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

19 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

20 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

21 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

22 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

23 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

24 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

25 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

26 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

27 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

28 (bases 1 to 222930)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 14:20:19 ; Search time 40 Seconds
(without alignments)
3141.790 Million cell updates/sec

Title: US-09-938-803-6
Perfect score: 487
Sequence: 1 MASSAEGDEGTVALAGVLQ.....ILPYHKAMVPRKELKME 487

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size: 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_protent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacterioph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	381	78.2	381	4 Q92980	Q92980 homo sapien
2	280	57.5	381	4 Q9BWA0	Q9BWA0 homo sapien
3	50	10.3	374	11 Q8BNH6	Q8BNH6 mus musculu
4	50	10.3	440	11 Q35131	Q35131 mus musculu
5	50	10.3	487	11 Q91V01	Q91V01 mus musculu
6	50	1.8	458	5 Q961A6	Q961A6 mus musculu
7	9	1.8	515	5 Q81P83	Q81P83 drosophila
8	9	1.8	556	5 Q9VL88	Q9VL88 drosophila
9	9	1.8	1462	10 Q9FIC5	Q9FIC5 arabidopsis
10	9	1.8	1486	10 Q9FXB7	Q9FXB7 arabidopsis
11	8	1.6	78	6 Q9GUX2	Q9GUX2 sus scrofa
12	8	1.6	86	6 Q8MN94	Q8MN94 oryctolagus
13	8	1.6	87	6 Q9TSG2	Q9TSG2 sus scrofa
14	8	1.6	87	13 Q8QFT5	Q8QFT5 gallus galli
15	8	1.6	299	10 Q8W0Y7	Q8W0Y7 medicago tr
16	8	1.6	541	17 Q97BR4	Q97BR4 thermoplasm

17	8	1.6	547	16 Q92Z88	Q92Z88 rhizobium m
18	8	1.6	588	10 Q9FIC0	Q9FIC0 arabidopsis
19	8	1.6	599	10 Q93VM9	Q93VM9 arabidopsis
20	8	1.6	599	10 Q8H134	Q8H134 arabidopsis
21	8	1.6	824	2 Q8VNZ4	Q8VNZ4 bordetella
22	8	1.6	1233	4 Q9UK61	Q9UK61 homo sapien
23	8	1.6	1854	5 Q9VTY8	Q9VTY8 drosophila
24	7	1.4	13	11 Q62354	Q62354 mus musculu
25	7	1.4	37	11 Q9OV33	Q9OV33 rattus sp.
26	7	1.4	54	16 Q92G83	Q92G83 rickettsia
27	7	1.4	55	16 Q970D5	Q970D5 streptococc
28	7	1.4	67	16 Q8NUZ2	Q8NUZ2 stephylecoc
29	7	1.4	82	10 Q93XT1	Q93XT1 diospyros h
30	7	1.4	91	2 Q9ACD7	Q9ACD7 vibrio salm
31	7	1.4	93	15 Q79559	Q79559 human immun
32	7	1.4	100	16 Q8CYQ8	Q8CYQ8 streptococc
33	7	1.4	101	3 Q07236	Q07236 saccharomyc
34	7	1.4	102	4 Q9B210	Q9B210 homo sapien
35	7	1.4	103	2 Q8RPN0	Q8RPN0 entilichia c
36	7	1.4	110	12 Q8VAV8	Q8VAV8 white spot
37	7	1.4	115	4 Q9UDH5	Q9UDH5 homo sapien
38	7	1.4	118	16 Q8F6V0	Q8F6V0 leptospira
39	7	1.4	129	17 Q8U237	Q8U237 pyrococcus
40	7	1.4	132	10 Q9SQ73	Q9SQ73 gossypium h
41	7	1.4	132	10 Q9SQ74	Q9SQ74 gossypium h
42	7	1.4	143	16 Q8ED53	Q8ED53 shewanella
43	7	1.4	150	17 Q9HSP8	Q9HSP8 halobacteri
44	7	1.4	154	4 Q9Y628	Q9Y628 homo sapien
45	7	1.4	157	2 Q8RR49	Q8RR49 streptococc

ALIGNMENTS

RESULT 1

ID	Q92980	PRELIMINARY;	PRT;	381 AA.
AC	Q92980;			
DT	01-FEB-1997 (TREMBLrel. 02, Created)			
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
GN	C3F.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Ansari-Lari M.A., Muzny D.M., Lu J., Lu F., Lilley C.E., Spanos S.,			
RL	Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96303695; PubMed=8723724;			
RA	Ansari-Lari M.A., Muzny D.M., Lu J., Lu F., Lilley C.E., Spanos S.,			
RA	Malley T., Gibbs R.A.;			
RT	"A gene-rich cluster between the CD4 and triosephosphate isomerase			
RT	genes at human chromosome 12p13."			
RL	Genome Res. 6:314-326(1996).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Ansari-Lari M.A., Muzny D.M., Lu J., Lu F., Lilley C.E., Spanos S.,			
RA	Malley T., Gibbs R.A.;			
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RA	Ansari-Lari M.A., Shen Y., Muzny D.M., Lee W., Gibbs R.A.;			
RA	Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RA	Ansari-Lari M.A., Shen Y., Muzny D.M., Lee W., Gibbs R.A.;			
RA	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.			
RL	EMBL; U72515; AAC51640.1; -			

DR EMBL; U47924; AAB51326.1; -
 DR InterPro; IPR001064; Crystallin.
 DR InterPro; IPR004299; MBOAT_fam.
 DR Pfam; PF03062; MBOAT_1.
 DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
 SQ SEQUENCE 381 AA; 4412 MW; F2AFA658155113 CRC64;

Query Match 78.2%; Score 381; DB 4; Length 381;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MGRITAVLTTCFQMAVYLLAGYYTATGNVDIKMTMPCVLTLLKLGAVDPGQKQ 166
 DB 1 MGRITAVLTTCFQMAVYLLAGYYTATGNVDIKMTMPCVLTLLKLGAVDPGQKQ 60
 QY 167 NSLSSEQOKYAIRGVPSLLEVAGFSYFGAPLVGPQFSNMHYMKLVQSELIDIPKIPNS 226
 DB 61 NSLSSEQOKYAIRGVPSLLEVAGFSYFGAPLVGPQFSNMHYMKLVQSELIDIPKIPNS 120
 QY 227 IIPALKRSLGLFYLVGYTLLSPHITEDYLLTEDYDNHFMFRCMYMLIWGKFLVYKYVT 286
 DB 121 IIPALKRSLGLFYLVGYTLLSPHITEDYLLTEDYDNHFMFRCMYMLIWGKFLVYKYVT 180
 QY 287 CWLVTEGVCIITGLGFNGFEKKGAKMDACANMKVWLFETNPRFTGTIASFNINNAWA 346
 DB 181 CWLVTEGVCIITGLGFNGFEKKGAKMDACANMKVWLFETNPRFTGTIASFNINNAWA 240
 QY 347 RYIFRKLKFLGNKELSGSLSLFLALMHGSHGYLVCFQMEFLIVVEQARLQESPT 406
 DB 241 RYIFRKLKFLGNKELSGSLSLFLALMHGSHGYLVCFQMEFLIVVEQARLQESPT 300
 QY 407 LSKLAITVLOPFYLVQTIHMLFMGYSMTAFCLFTMDKMLKVYSIYFLGHIFELSL 466
 DB 301 LSKLAITVLOPFYLVQTIHMLFMGYSMTAFCLFTMDKMLKVYSIYFLGHIFELSL 360
 QY 467 FILPYIHKAMVPRKEKLKME 487
 DB 361 FILPYIHKAMVPRKEKLKME 381

RESULT 2

Q9BM40 PRELIMINARY; PRT; 381 AA.
 AC Q9BM40;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 OS Homo sapiens (Human).
 OC Hypothetical protein.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RA Strausberg R.;
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; BC000664; AA000664.1; -
 DR InterPro; IPR001064; Crystallin.
 DR InterPro; IPR004299; MBOAT_fam.
 DR Pfam; PF03062; MBOAT_1.
 DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
 DR Hypothetical protein.
 SQ SEQUENCE 381 AA; 4412 MW; 2BF19AB389F39B1 CRC64;

Query Match 57.5%; Score 280; DB 4; Length 381;
 Best Local Similarity 100.0%; Pred. No. 1.3e-284;
 Matches 280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MGRITAVLTTCFQMAVYLLAGYYTATGNVDIKMTMPCVLTLLKLGAVDPGQKQ 166
 DB 1 MGRITAVLTTCFQMAVYLLAGYYTATGNVDIKMTMPCVLTLLKLGAVDPGQKQ 60

QY 167 NSLSSEQOKYAIRGVPSLLEVAGFSYFGAPLVGPQFSNMHYMKLVQSELIDIPKIPNS 226
 DB 61 NSLSSEQOKYAIRGVPSLLEVAGFSYFGAPLVGPQFSNMHYMKLVQSELIDIPKIPNS 120

QY 227 IIPALKRSLGLFYLVGYTLLSPHITEDYLLTEDYDNHFMFRCMYMLIWGKFLVYKYVT 286
 DB 121 IIPALKRSLGLFYLVGYTLLSPHITEDYLLTEDYDNHFMFRCMYMLIWGKFLVYKYVT 180

QY 287 CWLVTEGVCIITGLGFNGFEKKGAKMDACANMKVWLFETNPRFTGTIASFNINNAWA 346
 DB 181 CWLVTEGVCIITGLGFNGFEKKGAKMDACANMKVWLFETNPRFTGTIASFNINNAWA 240

QY 347 RYIFRKLKFLGNKELSGSLSLFLALMHGSHGYLVCFQMEFLIVVEQARLQESPT 406
 DB 241 RYIFRKLKFLGNKELSGSLSLFLALMHGSHGYLVCFQMEFLIVVEQARLQESPT 300

RESULT 3

O8BNH6 PRELIMINARY; PRT; 374 AA.
 AC O8BNH6;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Unknown Est (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Body;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK083687; BAC38993.1; -
 FT NON_TER 1
 SQ SEQUENCE 374 AA; 43259 MW; 68E4BD6B72A98486 CRC64;

Query Match 10.3%; Score 50; DB 11; Length 374;
 Best Local Similarity 100.0%; Pred. No. 2.5e-43;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 GTIASFNINNAWARYIFKRLKFLGNKELSGSLSLFLALMHGSHGYL 381
 DB 219 GTIASFNINNAWARYIFKRLKFLGNKELSGSLSLFLALMHGSHGYL 268

RESULT 4

O35131 PRELIMINARY; PRT; 440 AA.
 AC O35131;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE C3F (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=98112780; PubMed=9445485;
 RA Ansari-Lari M.A., Oeltyen J.C., Schwartz S., Zhang Z., Muzny D.M.,
 RA Lu J., Gorrell J.H., Chinault A.C., Belmont J.W., Miller W.,
 RA Gibbs R.A.;
 RT "Comparative sequence analysis of a gene-rich cluster at human
 RT chromosome 12p13 and its syntenic region in mouse chromosome 6.";
 RL Genome Res. 8:29-40(1998).
 RN [2]

```

RP SEQUENCE FROM N.A.
RA Muzny D., Ansari-Lari M.A., Timms K.M., Yu W., Dugan S., Lu J.,
RA Shen Y., Rowland K., Liu W., Perez L., Ding Y., Haywood M., Jain A.,
RA Leal B., Logan O., Nguyen V., Savage L., Shen H., Morley K., Chen E.,
RA Forcum J., Arenson A.D., Chiu M.W., Gorrell J.H., Brundage E., Di W.,
RA Chihault C., Nelson D., Gibbs R.A.;
RT "Direct Submission";
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Muzny D., Arenson A.D., Adams C., Brundage E., Bunac C., Carvelli K.,
RA Chacko J., Chen J., Di W., Ding Y., Dugan S., Durbin J., Forcum J.,
RA Ganesh R., Garcia C., Goodman M., Gorrell J.H., Haywood M.,
RA Hernandez J., Jackson L., Jin S., Kampal R., Karpachy S., Kovar C.,
RA Leal B., Li Y., Lichtarge O., Liu W., Logan O., Lu J., Ly T.,
RA Martinez C., Osval G., Perez L., Rashid N.D., Rowland K., Savage L.,
RA Scherer S.E., Shen H., Simon M., Stovall K., Timms K.M., Todd J.,
RA Vo Q., Williamson A., Worley K.C., Yu W., Chihault C., Nelson D.,
RA Gibbs R.A.;
RT "Direct Submission";
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AC002397; AAC36007.1;
DR InterPro: IPR004299; MBOAT_fam.
DR Pfam: PF03062; MBOAT_1.
DR NON_TER
SQ SEQUENCE 440 AA; 5117 MW; 7A78BAAB40A1C638 CRC64;

Query Match 10.3%; Score 50; DB 11; Length 440;
Best Local Similarity 100.0%; Pred. No. 2.9e-43; Mismatches 0; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 332 GTTASFNNTNAAWARYIFKRLKFLGNKELSGSLFLALMHLGHSYL 381
Db 285 GTTASFNNTNAAWARYIFKRLKFLGNKELSGSLFLALMHLGHSYL 334

RESULT 5
ID Q91V01 PRELIMINARY; PRT; 487 AA.
AC Q91V01;
DT 01-DEC-2001 (TRENBLREL. 19, Created)
DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Putative transmembrane protein PTG (Unknown) (Protein for
DE MGC:11670).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Liver;
RA Zhu Y., Han Y., Reddy J.K.;
RT "Cloning and initial characterization of mouse PTG cDNA, whose
RT expression is in a PPAR alpha dependent manner.";
RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY028317; AAK02915.1;
DR EMBL: BC006753; AA06753.2;
DR InterPro: IPR004299; MBOAT_fam.
DR Pfam: PF03062; MBOAT_1.
DR Transmembrane.
SQ SEQUENCE 487 AA; 56146 MW; EBC51DB3734B17C7 CRC64;

Query Match 10.3%; Score 50; DB 11; Length 487;
Best Local Similarity 100.0%; Pred. No. 3.1e-43; Mismatches 0; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 332 GTTASFNNTNAAWARYIFKRLKFLGNKELSGSLFLALMHLGHSYL 381
Db 332 GTTASFNNTNAAWARYIFKRLKFLGNKELSGSLFLALMHLGHSYL 381

RESULT 6
ID Q961A6 PRELIMINARY; PRT; 458 AA.
AC Q961A6;
DT 01-DEC-2001 (TRENBLREL. 19, Created)
DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
DE LD25907P (CG3779-PC).
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayan A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frie E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Munoz J., Paclebo J., Pargass V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Vandeil M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayan A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Baer A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavli S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin J.K., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kerchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclebo J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Skupski M.P., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodruff M., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Celisner S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,

```

RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreras S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleby J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svitskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
 RT "Sequencing of *Drosophila melanogaster* genome.",
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,
 RA Clamp M., Drysdale R., Marshall B., Milburn G., Richter J., Russo S.,
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
 RT "Annotation of *Drosophila melanogaster* genome.",
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Flybase; (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY051728; AAK93152.1; -
 DR EMBL; AB003625; AANI0693.1; -
 DR Flybase; FBgn0002973; numb.
 DR InterPro; IPR006020; PTB_PIB.
 DR Pfam; PF0640; PID; 1.
 DR SMART; SM00462; PTB; 1.
 DR PROSITE; PS01179; PID; 1.
 DR PROSITE; PS01179; PID; 1.
 SQ SEQUENCE 458 AA; 49315 MW; DF3FE1010233C3C2 CRC64;
 Query Match 1.8%; Score 9; DB 5; Length 458;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 360 ELSQGLSL 368
 Db 271 ELSQGLSL 279
 RESULT 7
 O8IPE9 PRELIMINARY; PRT; 515 AA.
 AC O8IPE9;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE CG3779-PB.
 GN NUMB.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton R.C., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.G., Rogers Y.H., Blazey R.C., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gaber G.L.,
 RA Abiri J.F., Adiyanti A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Balley R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Broctier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy J., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleby J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kimios I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weltschbach J.,
 RA Williams S.M., Woodgerl, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amaralides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreras S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleby J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svitskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
 RT "Sequencing of *Drosophila melanogaster* genome.",
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,
 RA Clamp M., Drysdale R., Marshall B., Milburn G., Richter J., Russo S.,
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
 RT "Annotation of *Drosophila melanogaster* genome.",
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Flybase; (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB003625; AANI0693.1; -
 SQ SEQUENCE 515 AA; 56078 MW; 135B5B5282F0E519 CRC64;
 Query Match 1.8%; Score 9; DB 5; Length 515;
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 360 ELSQGLSTL 368
 Db 328 ELSQGLSTL 336

RESULT 8

Q9VLB8 PRELIMINARY; PRT; 556 AA.
 ID Q9VLB8
 AC Q9VLB8
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE NUMB protein.
 GN NUMB OR CG3779.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OK NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amathides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champs M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abtil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.W., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.A., Boutin J., Brockett P., Brotter P.,
 RA Butts K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng X., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gairdian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalaal M., Kaulush F., Karpen G.H., Ke Z., Kennison J.D., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Liakso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtenski R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195 (2000).
 DR EMBL; AE003625; AAF5276.1; -;
 DR HSSP; P16554; 2NMB.
 DR FLYBase; FBgn0002973; numB.
 DR InterPro; IPR006020; PTB_PTD.
 DR Pfam; PF00640; PTD_1.
 DR SMART; SM00462; PTB; 1.
 DR PROSITE; PS01179; PTD; 1.
 SQ SEQUENCE 556 AA; 60614 MW; 91B2695B5DB8405 CRC64;

Query Match 1.8%; Score 9; DB 5; Length 556;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 360 ELSQGLSTL 368
 Db 369 ELSQGLSTL 377

RESULT 9

Q9FIC5 PRELIMINARY; PRT; 1462 AA.
 ID Q9FIC5
 AC Q9FIC5
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Retroelement pol polyprotein-like.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC Eustoidia II; Brassicales; Brassicaceae; Arabidopsis.
 OK NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=99156233; PubMed=10048488;
 RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
 RT Sequence features of the regions of 1,081,958 bp covered by seventeen
 RT physically assigned P1 and TAC clones.";
 RL DNA Res. 5:379-391 (1998).
 DR EMBL; AB016892; BAB10837.1; -;
 DR EMBL; AB010694; BAB10837.1; JOINED.
 DR InterPro; IPR005162; Retrotrans_gag.
 DR Pfam; PF03732; Retrotrans_gag; 1.
 DR Pfam; PF00663; rve; 1.
 SQ SEQUENCE 1462 AA; 163504 MW; DA9700D7BFCB900B CRC64;

Query Match 1.8%; Score 9; DB 10; Length 1462;
 Best Local Similarity 100.0%; Pred. No. 8.6;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 SIFUGYPPA 55
 Db 716 SIFUGYPPA 724

RESULT 10

Q9FXB7 PRELIMINARY; PRT; 1486 AA.
 ID Q9FXB7
 AC Q9FXB7
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Putative retroelement polyprotein.
 GN P25P12.89.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC Eustoidia II; Brassicales; Brassicaceae; Arabidopsis.
 OK NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Federpiehl N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
 RA Altafi H., Nguyen M., Lam B., Southwick A., Bel, Buehler E., Chin C.,
 RA Chlou J., Choi E., Dunn P., Gonzalez A., Huang B., Kim C., Koo T.,
 RA Lee J.M., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharay N.,
 RA Pham P., Sakano H., Schwartz J., Shinn P., Thayer A., Torimi M.,
 RA Vayenberg M., Walker M., Yu G., Ecker J., Theologis A., Davis R.W.,
 RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AC009323; AAG09097.1; -;
 DR InterPro; IPR005162; Retrotrans_gag.
 DR InterPro; IPR001584; Rve.

DR InterPro; IPR001878; Znf_CCHC.
 DR Pfam; PF03732; Retrotrans_gag; 1.
 DR Pfam; PF00665; rve; 1.
 DR Pfam; PF00098; zf-CCHC; 1.
 DR PRINTS; PR00939; C2HCZNFINGER.
 DR SMART; SM00343; Znf_C2HC; 1.
 DR PROSITE; PS50158; ZF_CCHC; 1.
 KW PolyProtein.
 SQ SEQUENCE 1486 AA; 165725 MW; CAF70FA04CEBA690 CRC64;

Query Match 1.8%; Score 9; DB 10; Length 1486;
 Best Local Similarity 100.0%; Pred. No. 8.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 SIFLGYPFA 55
 |||||
 Db 752 SIFLGYPFA 760

RESULT 11

09GUX2 PRELIMINARY; PRT; 78 AA.
 AC 09GUX2; 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Diazepam binding inhibitor (Fragment).
 GN DBI.

OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxId=9823;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Skeletal muscle;
 RA Davoli R., Fontanesi L., Zambonelli P., Bigi D., Milc J., Cagnazzo M.,
 RA Cenci V., Braglia S., Russo V.;
 RT "Isolation and mapping of expressed sequence tags from porcine
 RT skeletal muscle: a contribution to the genomic transcript map of this
 RT tissue";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ301367; CAC21172.1; -.
 DR HSSP; P07107; IACA.
 DR InterPro; IPR000582; Ac_coa_bind_prot.
 DR Pfam; PR00887; ACP; 1.
 DR PRINTS; PR00689; ACOABINDINGP.
 DR ProDom; PD002965; Ac_coa_bind_prot; 1.
 DR PROSITE; PS00880; ACP; 1.
 FT NON TER 78
 RN [1]
 SQ SEQUENCE 78 AA; 8807 MW; 55D1861C52EC6FAC CRC64;

Query Match 1.6%; Score 8; DB 6; Length 78;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 308 KKGAKWDA 315
 |||||
 Db 51 KKGAKWDA 58

RESULT 12

08WN94 PRELIMINARY; PRT; 86 AA.
 AC 08WN94; 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Acyl-CoA-binding protein (ACBP) (Diazepam binding inhibitor) (DBI)
 DE (Endozepine) (EP).
 GN DBI.

OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI_TaxId=9986;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Barnack N.H., Liu H., Qian Z., Bilderback T.;
 RT "Localization of an endogenous diazepam ligand, acyl CoA binding
 RT protein, to Muller cells in rabbit retina";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -! FUNCTION: BINDS MEDIUM- AND LONG-CHAIN ACYL-CoA ESTERS WITH VERY
 CC HIGH AFFINITY AND MAY FUNCTION AS AN INTRACELLULAR CARRIER OF
 CC ACYL-CoA ESTERS. IT IS ALSO ABLE TO DISPLACE DIAZEPAM FROM THE
 CC BENZODIAZEPINE (BZD) RECOGNITION SITE LOCATED ON THE GABA TYPE A
 CC RECEPTOR. IT IS THEREFORE POSSIBLE THAT THIS PROTEIN ALSO ACTS AS
 CC A NEUROPEPTIDE TO MODULATE THE ACTION OF THE GABA RECEPTOR (BY
 CC SIMILARITY).
 CC -! SUBUNIT: MONOMER (BY SIMILARITY).
 CC -! SIMILARITY: BELONGS TO THE ACP FAMILY.

CC EMBL; AF407578; AAK9808.2; -.
 DR InterPro; IPR000582; Ac_coa_bind_prot.
 DR Pfam; PF00887; ACP; 1.
 DR PRINTS; PR00689; ACOABINDINGP.
 DR ProDom; PD002965; Ac_coa_bind_prot; 1.
 DR PROSITE; PS00880; ACP; FALSE_NEG.
 KW Transport; Lipid-binding; Acetylation.
 FT INIT MET 0
 FT MOD RES 1 1 ACETYLATION (BY SIMILARITY).
 SQ SEQUENCE 86 AA; 9784 MW; F55EF76BB434A0F1 CRC64;

Query Match 1.6%; Score 8; DB 6; Length 86;
 Best Local Similarity 100.0%; Pred. No. 7.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 308 KKGAKWDA 315
 |||||
 Db 50 KKGAKWDA 57

RESULT 13

09TSG2 PRELIMINARY; PRT; 87 AA.
 AC 09TSG2; 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Endozepine.

OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxId=9823;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Back Fat;
 RA Ramlal N., Suzuki H., Karnuah A.B., Hamastima N.;
 RT "Cloning of pig endozepine gene";
 RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB019792; BAA34531.1; -.
 DR HSSP; P07107; IACA.
 DR InterPro; IPR000582; Ac_coa_bind_prot.
 DR Pfam; PR00887; ACP; 1.
 DR PRINTS; PR00689; ACOABINDINGP.
 DR ProDom; PD002965; Ac_coa_bind_prot; 1.
 DR PROSITE; PS00880; ACP; 1.
 SQ SEQUENCE 87 AA; 9896 MW; 3225B53E8307DE38 CRC64;

Query Match 1.6%; Score 8; DB 6; Length 87;
 Best Local Similarity 100.0%; Pred. No. 7.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 308 KKGAKWDA 315
 |||||
 Db 51 KKGAKWDA 58

RESULT 14

Search completed: November 21, 2003, 14:24:00
 Job time : 42 secs

Q8QFT5 PRELIMINARY; PRT; 87 AA.
 AC Q8QFT5;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Diazepam binding inhibitor.
 GN DBI;
 OS Gallus gallus (Chicken).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=broiler breeder; TISSUE=Liver;
 RA Baader M.;
 RT "Drug-induction of Hepatic Cytochromes P450: Studies in Chicken
 Hepatoma Cells."
 RL Thesis (2002), Department of Biozentrum Pharmacology/Neurobiology,
 University of Basel, Basel, Switzerland.
 DR EMBL; AJ430511; CAD33129.1; -
 DR InterPro; IPR000582; AC_coh_bind_prot.
 DR Pfam; PF00887; ACBP; 1.
 DR ProDom; PD002965; AC_coh_bind_prot; 1.
 SQ SEQUENCE 87 AA; 9805 MW; 67405B1C62AB066 CRC64;

Query Match 1.6%; Score 8; DB 13; Length 87;
 Best Local Similarity 100.0%; Pred. No. 7.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 KGKAKMDA 315
 |||||
 Db 51 KGKAKMDA 58

RESULT 15
 Q8W0Y7 PRELIMINARY; PRT; 299 AA.
 AC Q8W0Y7;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Enod8.3 (Fragment).
 GN ENOD8.3.
 OS Medicago truncatula (Barrel medic).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
 OX NCBI_TaxID=3880;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Jemalong;
 RA Dickstein R., Hu X., Yang J., Ba L., Coque L., Kim D.-J., Cook D.R.,
 RA Yeung A.T.;
 RT "Randomly duplicated Enod8 genes are differentially expressed in
 Medicago."
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF463407; AAL68830.1; -
 DR InterPro; IPR001087; Lipase_GDSL.
 DR Pfam; PF00657; Lipase_GDSL; 1.
 FT NON_TER
 SQ SEQUENCE 299 AA; 33075 MW; F2A7D59A499471AB CRC64;

Query Match 1.6%; Score 8; DB 10; Length 299;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 KIPNSIIP 229
 |||||
 Db 33 KIPNSIIP 40

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 14:11:33 / Search time 39 seconds
(without alignments)
3222.349 Million cell updates/sec

Title: US-09-938-803-6
Perfect score: 2580
Sequence: 1 MASSAEGDEGVVALAGVLQ.....ILPYIKRAMPYRKELKME 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_protist:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacterioph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2314	89.7	487	11 Q91V01	Q91V01 mus musculu
2	2174	84.3	440	11 Q35131	Q35131 mus musculu
3	2046	79.3	381	4 Q92980	Q92980 homo sapien
4	2042	79.1	381	4 Q9BM40	Q9BM40 homo sapien
5	1856	71.9	374	11 Q8BNH6	Q8BNH6 mus musculu
6	874.5	33.9	497	5 Q9XVY9	Q9XVY9 mus musculu
7	871.5	33.8	497	5 Q9VXK5	Q9VXK5 drosophila
8	766	29.7	473	5 Q9VXK5	Q9VXK5 drosophila
9	408.5	15.8	722	5 Q9V5F2	Q9V5F2 caenorhabdi
10	407	15.8	465	3 Q9CAN8	Q9CAN8 arabidopsis
11	406	15.7	509	3 Q42916	Q42916 schizosacch
12	406	15.7	619	3 Q08548	Q08548 saccharomyc
13	383.5	14.9	462	10 Q8RWT4	Q8RWT4 arabidopsis
14	383.5	14.9	480	5 Q18856	Q18856 caenorhabdi
15	382.5	14.8	457	10 Q9LNB3	Q9LNB3 arabidopsis
16	373.5	14.5	519	11 Q8R312	Q8R312 mus musculu

17	373	14.5	492	11 Q8BRH8	Q8BRH8 mus musculu
18	372	14.4	492	11 Q8R3T2	Q8R3T2 mus musculu
19	344.5	13.4	487	11 Q9CZ73	Q9CZ73 mus musculu
20	341.5	13.2	487	11 Q8R192	Q8R192 mus musculu
21	300	11.6	386	11 Q8BRH5	Q8BRH5 mus musculu
22	297	11.5	540	5 Q17826	Q17826 caenorhabdi
23	280	10.9	378	11 Q8BWS6	Q8BWS6 mus musculu
24	266	10.3	350	4 Q96KY4	Q96KY4 homo sapien
25	263.5	10.2	451	5 Q8T8Q2	Q8T8Q2 drosophila
26	260.5	10.1	473	11 Q9C776	Q9C776 mus musculu
27	260	10.1	350	3 Q03130	Q03130 saccharomyc
28	255.5	9.9	473	11 Q8R1P9	Q8R1P9 mus musculu
29	254.5	9.9	472	4 Q96N66	Q96N66 homo sapien
30	254.5	9.9	473	11 Q8CHK3	Q8CHK3 mus musculu
31	244.5	9.5	480	5 Q19468	Q19468 caenorhabdi
32	242	9.4	303	4 Q8N9R5	Q8N9R5 homo sapien
33	241.5	9.4	489	5 Q9VMD5	Q9VMD5 drosophila
34	213.5	8.3	450	13 Q91938	Q91938 xenopus lae
35	212	8.2	343	4 Q9BVP2	Q9BVP2 homo sapien
36	212	8.2	455	13 Q91936	Q91936 xenopus lae
37	210.5	8.2	456	13 Q91937	Q91937 xenopus lae
38	208	8.1	461	13 Q91935	Q91935 xenopus lae
39	190	7.4	450	4 Q9H234	Q9H234 homo sapien
40	188.5	7.3	344	4 Q9BRE9	Q9BRE9 homo sapien
41	185	7.2	450	11 Q9J3J9	Q9J3J9 mus musculu
42	185	7.2	456	11 Q9J3K0	Q9J3K0 mus musculu
43	184	7.1	456	4 Q9H236	Q9H236 homo sapien
44	183.5	7.1	455	11 Q9J3J8	Q9J3J8 mus musculu
45	180.5	7.0	461	11 Q9J3J7	Q9J3J7 mus musculu

ALIGNMENTS

RESULT 1

ID	Q91V01	PRELIMINARY;	PRT;	487 AA.
AC	Q91V01;			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Putative transmembrane protein P7G (Unknown) (Protein for MGC:11670).			
DE	MGC:11670.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N. A.			
RC	STRAIN=C57BL/6J; TISSUE=Liver;			
RT	Zhu Y., Han Y., Reddy J.K.;			
RT	"Cloning and initial characterization of mouse P7G cDNA, whose expression is in a PPAR alpha dependent manner."			
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N. A.			
RA	Strauberg R.;			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AY028317; AK20915.1; -			
DR	EMBL; BC006753; AA06753.2; -			
DR	InterPro; IPR004299; MBOAT_fam.			
DR	Pfam; PF03062; MBOAT; 1.			
KW	Transmembrane.			
SO	SEQUENCE 487 AA; 56146 MW; EBC51DB3734B17C7 CRC64;			

Query Match	89.7%; Score 2314; DB 11; Length 487;
Best Local Similarity	87.3%; Pred. No. 1.1e-18;
Matches	425; Conservative 30; Mismatches 32; Indels 0; Gaps 0;
Qy	1 MASSAEGDEGVVALAGVLQSGFOELSLNKLATSLGASEQALRIITIFLCYPPALFYRH 60
Db	1 MASTADDMGTETOMRGLWPGVDLSLNKLATSLGASEQALRIITIFLCYPPALFYRH 60

```

QY 61 YLFYKETYLIHLEHFTFTGSIAYFNGNOLYHSLCTIVQFLIRLMGRTITAVLTTCF 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 YLFYKDSYLIHLEHFTFTGSIAYFNGNOLYHSLCTIVQFLIRLMGRTITAVLTTCF 120
QY 121 QMAVYLLAGYATGATNYDIKMTMPHCVLTLLKLGAVDYFDGKONSISSEQKYAIRG 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 QMAVYLLAGYATGATNYDIKMTMPHCVLTLLKLGAVDYFDGKONSISSEQKYAIRG 180
QY 181 VESLLEVGAFSYFYGAFLVGPFSNMHMKLVQGLIDIPGKIPNSIIPALRSLGLFY 240
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 VESLLEVGAFSYFYGAFLVGPFSNMHMKLVQGLIDIPGKIPNSIIPALRSLGLFY 240
QY 241 LVGYTLSPHITEDYLLTEDYDNHPPFRCMMLIGKFLVLYKYVTCMLVTEGVCILTL 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 LVGYTLSPHITEDYLLTEDYDNHPPFRCMMLIGKFLVLYKYVTCMLVTEGVCILTL 300
QY 301 GNGEPEKAKADACANMKVLFETNPRFTGTIASFNINTAMVARIYFKELKFLGNKE 360
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 GNGEPEKAKADACANMKVLFETNPRFTGTIASFNINTAMVARIYFKELKFLGNKE 360
QY 361 LSQGLSLFLALMHGHSGLVLCQFMEFLIVYERQAARLIOESPITSLAITVLQPY 420
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 LSQGLSLFLALMHGHSGLVLCQFMEFLIVYERQAARLIOESPITSLAITVLQPY 420
QY 421 YLVQQTIIHMLFMGYSTAFCLFTWDKMLKYKSIYFLGHIFFLSLFLPIYHKAMVPRK 480
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 YLVQQTIIHMLFMGYSTAFCLFTWDKMLKYKSIYFLGHIFFLSLFLPIYHKAMVPRK 480
QY 481 EKLKKME 487
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 481 EKLKKRE 487
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 2

035131 PRELIMINARY; PRT; 440 AA.

AC 035131; 01-JAN-1998 (TReMBLrel. 05, Created)

DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE C3F (Fragment).

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCB1_TaxID=10090;

RN [1] SEQUENCE FROM N.A.

RP MEDLINE=98112780; PubMed=9445485;

RX Ansari-Lari M.A., Oeltjen J.C., Schwartz S., Zhang Z., Muzny D.M., Lu J., Gorrell J.H., Chinault A.C., Belmont J.W., Miller W., Gibbs R.A.; "Comparative sequence analysis of a gene-rich cluster at human chromosome 12p13 and its syntenic region in mouse chromosome 6.";

RT Genome Res. 8:29-40(1998).

RL [2] SEQUENCE FROM N.A.

RN Shen Y., Ansari-Lari M.A., Timms K.M., Yu W., Dugan S., Lu J., Shen Y., Rowland K., Liu W., Perez L., Ding Y., Haywood M., Jain A., Leal B., Logan O., Nguyen V., Savage L., Shen H., Worley K., Chen E., Forcum J., Arenson A.D., Chiu M.W., Gorrell J.H., Brundage E., Di W., Chinault C., Nelson D., Gibbs R.A.; "Direct Submission.";

RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

RN [3] SEQUENCE FROM N.A.

RP Muzny D., Arenson A.D., Adams C., Brundage E., Bunac C., Carvelli K., Chacko J., Chen J., Di W., Ding Y., Dugan S., Durbin J., Forcum J., Ganeesh R., Garcia C., Goodman M., Gorrell J.H., Haywood M., Hernandez J., Jackson L., Jin S., Kampel R., Karpathy S., Kovar C., Leal B., Li Y., Lichtege O., Liu W., Logan O., Lu J., Ly T., Martinez C., Owal G., Perez L., Rashid N.D., Rowland K., Savage L., Scherer S.E., Shen H., Simon M., Scovall K., Timms K.M., Todd J., Vo O., Williamson A., Worley K.C., Yu W., Chinault C., Nelson D.,

```

RA Gibbs R.A.;
RT "Direct Submission.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC002397; AAC36007.1; -.
DR InterPro; IPR004299; MBOAT_fam.
DR Pfam; PF03062; MBOAT; 1.
FT NON TER
SQ SEQUENCE 440 AA; 5117 MW; 7A78BA840A1C638 CRC64;
Query Match 84.3%; Score 2174; DB 11; Length 440;
Best Local Similarity 89.7%; Pred. No. 8.2e-177;
Matches 393; Conservative 25; Mismatches 20; Indels 0; Gaps 0;
QY 50 LGYPALPYRYHLEPKETYLHLEHFTFTGSIAYFNGNOLYHSLCTIVQFLIRLMGR 109
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3 LGYPALPYRYHLEPKETYLHLEHFTFTGSIAYFNGNOLYHSLCTIVQFLIRLMGR 62
QY 110 TITAVLTTCFQOMAYLLAGYATGATNYDIKMTMPHCVLTLLKLGAVDYFDGKONS 169
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 63 TITAVLTTCFQOMAYLLAGYATGATNYDIKMTMPHCVLTLLKLGAVDYFDGKONS 122
QY 170 SSEQKYAIRGVPSLLEVAGFSYFYGAFLVGPFSNMHMKLVQGLIDIPGKIPNSIIP 229
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 123 SSEQKYAIRGVPSLLEVAGFSYFYGAFLVGPFSNMHMKLVQGLIDIPGKIPNSIIP 182
QY 220 ALKRLSLGLFYLVGYTLSPHITEDYLLTEDYDNHPPFRCMMLIGKFLVLYKYVTCML 289
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 183 ALKRLSLGLFYLVGYTLSPHITEDYLLTEDYDNHPPFRCMMLIGKFLVLYKYVTCML 242
QY 290 VTEGVCILTLGLFNGEPEKAKADACANMKVLFETNPRFTGTIASFNINTAMVARIY 349
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 243 VTEGVCILTLGLFNGEPEKAKADACANMKVLFETNPRFTGTIASFNINTAMVARIY 302
QY 350 FKRLLFLNGKLSQSLSLFLALMHGHSGLVLCQFMEFLIVYERQAARLIOESPITSL 409
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 303 FKRLLFLNGKLSQSLSLFLALMHGHSGLVLCQFMEFLIVYERQAARLIOESPITSL 362
QY 410 LAAIVLQPFYLVQQTIIHMLFMGYSTAFCLFTWDKMLKYKSIYFLGHIFFLSLFL 469
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 363 LAAIVLQPFYLVQQTIIHMLFMGYSTAFCLFTWDKMLKYKSIYFLGHIFFLSLFL 422
QY 470 PYHKAMVPRKELKME 487
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 423 PYHKAMVPRKELKRE 440
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 3

092980 PRELIMINARY; PRT; 381 AA.

AC 092980; 01-FEB-1997 (TReMBLrel. 02, Created)

DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE C3F.

GN GN

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.

OX NCB1_TaxID=9606;

RN [1] SEQUENCE FROM N.A.

RP Ansari-Lari M.A., Shen Y., Muzny D.M., Lee W., Gibbs R.A.;

RT Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

RN [2] SEQUENCE FROM N.A.

RP MEDLINE=96303695; PubMed=8723724;

RX Ansari-Lari M.A., Muzny D.M., Lu J., Lu F., Lilley C.E., Spanos S., Malley T., Gibbs R.A.;

RT "A gene-rich cluster between the CD4 and triosephosphate isomerase genes at human chromosome 12p13.";

RL Genome Res. 6:314-326(1996).

RN [3] SEQUENCE FROM N.A.

RA Anseri-Lari M.A., Muzny D.M., Lu J., Lu F., Lilley C.E., Spanos S.,
 RA Mailey T., Gibbs R.A.;
 RA Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RP Anseri-Lari M.A., Shen Y., Muzny D.M., Lee W., Gibbs R.A.;
 RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RP Anseri-Lari M.A., Shen Y., Muzny D.M., Lee W., Gibbs R.A.;
 RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL; U72515; AAC13640.1; -
 DR EMBL; U47924; AAB1326.1; -
 DR InterPro; IPR001064; Crystalin.
 DR InterPro; IPR004239; MBOAT_fam.
 DR Pfam; PF03062; MBOAT_1.
 DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
 SQ SEQUENCE 381 AA; 44122 MW; FD2AF658155113 CRC64;

Query Match 79.3%; Score 2046; DB 4; Length 381;
 Best Local Similarity 100.0%; Pred. No. 5.5e-166;
 Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 MGRITAVLTTCFQMAVLLAGYTTATGNVDIKWTMPCVLTLLKLGAVDPFGKQD 166
 DB 1 MGRITAVLTTCFQMAVLLAGYTTATGNVDIKWTMPCVLTLLKLGAVDPFGKQD 60
 QY 167 NSLSEQQKXAIRGVPSLLEVAQSFYGAFLVGPQFSNNHYMKLVQGLIDIPKIPNS 226
 DB 61 NSLSEQQKXAIRGVPSLLEVAQSFYGAFLVGPQFSNNHYMKLVQGLIDIPKIPNS 120
 QY 227 IIPALKRSLGLFYLVGTLLSPHITEDYLLTEDYDNPFWRCMYMLIWGKFLVKYVT 286
 DB 121 IIPALKRSLGLFYLVGTLLSPHITEDYLLTEDYDNPFWRCMYMLIWGKFLVKYVT 180
 QY 287 CMLVTEGVCITLGLGNGEPEEKAKWDACANMKVLFETNPRFTGTIASFNINTNAWA 346
 DB 181 CMLVTEGVCITLGLGNGEPEEKAKWDACANMKVLFETNPRFTGTIASFNINTNAWA 240
 QY 347 RYIFRKLFLGNKELSGSLFLALMHGSHGYLVCFQMEFLIYVERQARLIQESPT 406
 DB 241 RYIFRKLFLGNKELSGSLFLALMHGSHGYLVCFQMEFLIYVERQARLIQESPT 300
 QY 407 LSKLAITVLOPFYLVQOTIHWLFMGYSMTAFCLFTWDKMLKVYSIYFLGHIFLSLL 466
 DB 301 LSKLAITVLOPFYLVQOTIHWLFMGYSMTAFCLFTWDKMLKVYSIYFLGHIFLSLL 360
 QY 467 FILPIYHKAMVPRKEXLKKME 487
 DB 361 FILPIYHKAMVPRKEXLKKME 381

RESULT 4

Q9BNW40 PRELIMINARY; PRT; 381 AA.
 ID Q9BNW40;
 AC Q9BNW40;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Colon;
 RA Steniusberg R.;
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; BC000664; AAH00664.1; -
 DR InterPro; IPR001064; Crystalin.
 DR InterPro; IPR004239; MBOAT_fam.
 DR Pfam; PF03062; MBOAT_1.

DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 381 AA; 44121 MW; 2BF19AB369FF39B1 CRC64;

Query Match 79.1%; Score 2042; DB 4; Length 381;
 Best Local Similarity 99.7%; Pred. No. 1.2e-165;
 Matches 380; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 107 MGRITAVLTTCFQMAVLLAGYTTATGNVDIKWTMPCVLTLLKLGAVDPFGKQD 166
 DB 1 MGRITAVLTTCFQMAVLLAGYTTATGNVDIKWTMPCVLTLLKLGAVDPFGKQD 60
 QY 167 NSLSEQQKXAIRGVPSLLEVAQSFYGAFLVGPQFSNNHYMKLVQGLIDIPKIPNS 226
 DB 61 NSLSEQQKXAIRGVPSLLEVAQSFYGAFLVGPQFSNNHYMKLVQGLIDIPKIPNS 120
 QY 227 IIPALKRSLGLFYLVGTLLSPHITEDYLLTEDYDNPFWRCMYMLIWGKFLVKYVT 286
 DB 121 IIPALKRSLGLFYLVGTLLSPHITEDYLLTEDYDNPFWRCMYMLIWGKFLVKYVT 180
 QY 287 CMLVTEGVCITLGLGNGEPEEKAKWDACANMKVLFETNPRFTGTIASFNINTNAWA 346
 DB 181 CMLVTEGVCITLGLGNGEPEEKAKWDACANMKVLFETNPRFTGTIASFNINTNAWA 240
 QY 347 RYIFRKLFLGNKELSGSLFLALMHGSHGYLVCFQMEFLIYVERQARLIQESPT 406
 DB 241 RYIFRKLFLGNKELSGSLFLALMHGSHGYLVCFQMEFLIYVERQARLIQESPT 300
 QY 407 LSKLAITVLOPFYLVQOTIHWLFMGYSMTAFCLFTWDKMLKVYSIYFLGHIFLSLL 466
 DB 301 LSKLAITVLOPFYLVQOTIHWLFMGYSMTAFCLFTWDKMLKVYSIYFLGHIFLSLL 360
 QY 467 FILPIYHKAMVPRKEXLKKME 487
 DB 361 FILPIYHKAMVPRKEXLKKME 381

RESULT 5

Q9BNW6 PRELIMINARY; PRT; 374 AA.
 ID Q9BNW6;
 AC Q9BNW6;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Unknown EST (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Body;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium;
 RT "Analysis of the mouse transcription Research Group Phase I & II Team;
 RL Nature 420:565-573 (2002)."
 DR EMBL; AK083687; BAC38993.1; -
 FT NON_TER
 SQ SEQUENCE 374 AA; 43259 MW; 68E4BD6B72A98486 CRC64;

Query Match 71.9%; Score 1856; DB 11; Length 374;
 Best Local Similarity 89.8%; Pred. No. 8.1e-150;
 Matches 336; Conservative 20; Mismatches 18; Indels 0; Gaps 0;

QY 114 VLTTCFQMAVLLAGYTTATGNVDIKWTMPCVLTLLKLGAVDPFGKQDQNSLSEQ 173
 DB 1 VLTTCFQMAVLLAGYTTATGNVDIKWTMPCVLTLLKLGAVDPFGKQDQNSLSEQ 60
 QY 174 OKYAIRGVPSLLEVAQSFYGAFLVGPQFSNNHYMKLVQGLIDIPKIPNSIIPALKR 233
 DB 61 OKYAIRGVPSLLEVAQSFYGAFLVGPQFSNNHYMKLVQGLIDIPKIPNSIIPALKR 120

```

QY 234 LSLGLFVLYGVYTLSPHITTEDEYLLTEYDNDHPFPCOMWLIMGKFLVLYKYVTGMLVTBG 293
DB 121 LSLGLFVLYGVYTLSPHITTEDEYLLTEYDNDHPFPCOMWLIMGKFLVLYKYVTGMLVTBG 180
QY 294 VCILITGLFNGEPEEKAKADACANNKWLFEETNPFTGTIASFNINTANAVARIYFKRL 353
DB 181 VCILITGLFNGEPEEKAKADACANNKWLFEETNPFTGTIASFNINTANAVARIYFKRL 240
QY 354 KFLGNELISOGSLTLFLALMHGHSGLVCFQMEFLIVYKQVSSLIRDSPLSLSLAI 413
DB 241 KFLGNELISOGSLTLFLALMHGHSGLVCFQMEFLIVYKQVSSLIRDSPLSLSLAI 300
QY 414 TVLQPFYVLYVQOQTIHMLFMGYSMTAFCLFTWDMKLKYKSIYFLGHIFLSTLFLPYTH 473
DB 301 TALQPFYVLYVQOQTIHMLFMGYSMTAFCLFTWDMKLKYKSIYFLGHIFLSTLFLPYTH 360
QY 474 KAMVPRKREKLKQME 487
DB 361 KAMVPRKREKLKQME 374

```

RESULT 6

```

Q9XYV9 PRELIMINARY; PRT; 497 AA.
ID 09XYV9
AC 09XYV9;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Putative transmembrane protein NES5.
DE NES OR NES5 OR CG9655.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN 1[1]
RP SEQUENCE FROM N.A.
RA Mauriell-Zaffran C., Chauvet S., Miasod R., Jullien N., Pradel J.,
RA Aragnol D.,
RT "The evolutionary conserved gene nes5 is a putative target of Hox
RT proteins in Drosophila embryo."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF28112; AAD28257.1; -
DR FlyBase: FBgn0026630; nes-
DR InterPro: IPR004299; MBOAT_fam.
DR Pfam: PF03062; MBOAT; 1.
KW Transmembrane.
SQ SEQUENCE 497 AA; 57436 MW; FBF535BI3700520B CRC64;

```

```

Query Match 33.9%; Score 874.5; DB 5; Length 497;
Best Local Similarity 37.2%; Pred. No. 4.4e-66;
Matches 167; Conservative 92; Mismatches 165; Indels 25; Gaps 5;
QY 28 LNKLTSLGASQALRLISIFLPYRHYL-FYKETYLIHLFHTFTGLSIAYFNF 86
DB 14 MDGIASGVGVPAALRLILITLILAGYPAALYQKFIIVADKYVHHNFRCAGGLCYFNF 73
QY 87 GNOLYHSLCTV-LQFLILRLKRTTAVLTTCFQMAVYLAGYYTATGNYDKMTMPH 145
DB 74 GRDTYHSLALITTYFLVILRLKKTQFLAIVNFVHMSYLLLCGYFTSSNDYDILMTMPH 133
QY 146 CVTLKILGAVVYPPGQKQNSLSQOKAIVRQVPSLLEVAGSFYAFIVGPOFSM 205
DB 134 CILVLNMGIGFDITDGLKESELSKQKTAALKKPSLELAFSTFPGFLVGFQFPP 193
QY 206 NHYMKLVQGEILIDIPKIPMSIIPALKRLSLGLFVLYGVYTLSPHITTEDEYDNDHP 265
DB 194 RRRKAFVDFGEFRHQHEGVANGV----RRFGAGFYLLVQVGYGRYLPDSYFLTPPEFAQVS 249
QY 266 FWFRCMYMLIWGFVLYKYVTGMLVTGVCILITGLFNGEPEEKAKADACANNKWLFE 325
DB 250 FVRIYLLGFQWAKFSLYKYISWLLTEGALICIGLTFYKGBDKNGQPDWSCSNVKLLE 309

```

```

QY 326 TNPRTGTIASFNINTANAVARIYFKRLKFLGNELISOGSLTLFLALMHGHSGLVCFQ 385
DB 310 TGMTNHHVYQSPNVTNMGVGYIYKRLKFLNNRRTISGALGFLALVHGHYSGYMTL 369
QY 386 MEFLIVYKQVSSLIRDSPLSLSLAI TVLQPFYVLYVQOQTIHMLFMGYSMTAFCLFTWDMKLKYKSIYFLGHIFL 463
DB 370 MEYVAVSTREKQITRFYTKVFLPQWGHILNNSDIYKL-----LYFITLKSYNVVMGM 421
QY 435 SMTAFCLFTWDMKLKYKSIYFLGHIFL 463
DB 422 CLTAFLVFLKERYIVVYGAVSYYGFTFLV 450

```

RESULT 7

```

Q9VYX5 PRELIMINARY; PRT; 497 AA.
ID 09VYX5
AC 09VYX5;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE NES protein (R03440P).
DE NES OR CG9655.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN 1[1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20196006; PubMed=10731132;
RX STRAIN=BERKELEY;
RC STRAIN=BERKELEY;
RA Adams M.D., Gelinkner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amarantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhendari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodali C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy B., Murzyn D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusseken D.R., Pacleb J.M.,
RA Palazonio M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA Reimert K., Remington K., Saunders R.D.C., Schaefer F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskae R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassatman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN 1[2]
RP SEQUENCE FROM N.A.
RX STRAIN=berkeley;

```

RA Stapleton M., Brokstein P., Hong L., Abmayari A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Paclet J., Paragass V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celinker S.,
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AEO03517; AAF49181.1; -
 DR EMBL: AY070936; AAL48558.1; -
 DR FlyBase: FBgn0026630; nes
 DR InterPro: IPR004299; MBOAT_fam.
 DR Pfam: PF03062; MBOAT_1.
 DR SEQUENCE 497 AA; 57393 MW; 4927EFBD784BF861 CRC64;

Query Match 33.8%; Score 871.5; DB 5; Length 497;
 Best Local Similarity 37.2%; Pred. No. 7.8e-66;
 Matches 167; Conservative 92; Mismatches 165; Indels 25; Gaps 5;

QY 28 LNKLTSLGASGQALRLIISIFLGYPPALFYRHYL-FYKETYLIHLPTFTGSLAYFNF 86
 DB 14 MDGIASGVVPEVLELRLITLILAGYPAALYOKFISVADKTVHMFAGGAGLGYFNY 73
 QY 87 GNQLYHSLCLIV-LQFLILRLMGRTTAVLTTCFQMAVYLAGYVYATGNYDIKTMPH 145
 DB 74 GLDTYHSLILTLTYFLVLRKKTQIFLAIFVFNHMSYLLIGYFTSSNDYDILMTMPH 133
 QY 146 CVLTLKTLGLAVDFDGDKNQSLSSBOQKAIKRGVSLLEVAAGSYFYGAFLVNGQPSM 205
 DB 134 CILVLRNIGGFDITDGLKESELSKQOKTALKKPSLELLAFSPSGFLVGPQPP 193
 QY 206 NHYMKLVQGLIDIPKIPNSIIPALKRLSLGLFVLYGYTLSPHITEDYLLTDDYNH 265
 DB 194 RRYAFVDFGEHREGHVEAGV---RRFGAFYLLVCGVGLAYLDSYFLTEPQVS 249
 QY 266 FWFRCMTMLWGKVLKYVTCMLVTEGVCITLGLGFNGFEKAKADACANKWMLFE 325
 DB 250 FVKRIYLLGFMAKPSLYKYISCMILTGGALICILYKGBDKNQCPMWSGCSNVKXLE 309
 QY 326 TNPFTTASFNNTAMAVARYIFKRLKPLGNKELSGSLFLALMHGHSGLVCFQ 385
 DB 310 TGNTEHYVOSFNNTOMQOYIYKRLKFLNNTISGALGLVAWHGHSGLYVTF 369
 QY 386 MEFLIVEROPARL-----IQESPFLSKLAITVLOPFYVLYVQOTIHLFMGY 434
 DB 370 MEVWVSTEQITFTYRKVVLPMQGHILNMSDIYKL-----LYFTLTKSVVVMGW 421
 QY 435 SMTAFCLFTMDKWLKVTKSYIFLGHIFPL 463
 DB 422 CLTAFVFLKXERWIVVGVAVSYGFTFLV 450

RESULT 8
 ID 001925 PRELIMINARY; PRT; 473 AA.
 AC 001925;
 DT 01-JUL-1997 (Tremblrel. 04, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Hypothetical protein R155.1.
 GN R155.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitidae;
 OC Rhabdilitidae; Peloderiinae; Caenorhabditis.
 OX NCB1_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA Waterston R.;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Gaisel C., Mameley P., Kramer J.;
 RT "The sequence of C. elegans cosmid R155."
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF003390; AAB54271.2; -
 KM Hypothetical protein.
 KM SEQUENCE 473 AA; 54113 MW; 2198AA79247D1A77 CRC64;

Query Match 29.7%; Score 766; DB 5; Length 473;
 Best Local Similarity 33.8%; Pred. No. 7.1e-57;
 Matches 157; Conservative 95; Mismatches 203; Indels 10; Gaps 3;

QY 31 LATSLSGSEQALRLIISIFLGYPPALFYRHYL-FYKETYLIHLPTFTGSLAYFNF 90
 DB 7 LSEVTSASEDALRLISVLAGYPAALYVHRFTFNKRAQHOHLFFVIGSLMMFNCGSSV 66
 QY 91 YHSLCLIVLQFLILRLMGRTTAVLTTCFQMAVYLAGYVYATGNYDIKTMPHCYTL 150
 DB 67 IHPILSIFGAFYTNFMAAGTASIVAAHIVFLGLHLIGVWFHETDYDITWTTPFCIMTL 126
 QY 151 KLGLAVDFDGDKNQSLSSBOQKAIKRGVSLLEVAAGSYFYGAFLVNGQPSM 210
 DB 127 RFLGLVNDVVDGQKPEHLPDQKLTIRISPKGLLETAAGLFFQGLTVLPQFTLSKFRS 186
 QY 211 LVQGLIDIPKIPNS-IIPALKRLSLGLFVLYGYTLSPHITEDYLLTDDYNH 269
 DB 187 FVNGDMLDSDGQPKKAFPLPSISGFLAGCTYVWLHQGQWIDQYNSDANNLSFMR 246
 QY 270 CMTMLWGKVLKYVTCMLVTEGVCITLGLGFNGFEKAKADACANKWMLFE 329
 DB 247 WSWVTLFMFLTWKYCAMWLTGASLISGLHNGKABENDWDGGRDHIIMKMGCHD 306
 QY 330 FTGTIASFNNTAMAVARYIFKRLKPLGNKELSGSLFLALMHGHSGLVCFQ 389
 DB 307 YNSVESFNGCTVTPFAKHILHRRIRWNNKLASHVITLSTLALMHGHSGLVCFQ 366
 QY 390 IIVVEROAPALIQESPFLSKLAITVLOPFYVLYVQOTIHLFMGYSTACFLFTMDKWLK 449
 DB 367 CVAQNLQVLLIKRTPMSAISKPSRPFIMFGKLTISYNGAFALMGLITKTMIG 426
 QY 450 VVNSIYFLGHIFP-----LSLFTLP-YIHKAVPRKELKX 485
 DB 427 PVKSIVFIFGIYFIVWPIILHMLRLVLPHPKKAALAEKEBEVKK 471

RESULT 9
 ID 09V5F2 PRELIMINARY; PRT; 722 AA.
 AC 09V5F2;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE CG18445 protein.
 GN CG18445.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Prexygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCB1_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherter S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baker E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abin J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Broctier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu L.B., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palczolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtka R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Messarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.W., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA "The genome sequence of *Drosophila melanogaster*.";
 RA Science 287:2185-2195(2000).
 DR EMBL: AE003831; AAF58858.1;
 DR Flybase: FBgn003476; CG18445.
 DR InterPro: IPR004299; MBOAT_fam.
 DR Pfam: PF03062; MBOAT; 1.
 SQ SEQUENCE 722 AA; 81221 MW; CBAE84FA6720E367 CRC64;

Query Match 15.8%; Score 408.5; DB 5; Length 722;
 Best Local Similarity 25.8%; Pred. No. 3.3e-26;
 Matches 121; Conservative 89; Mismatches 212; Indels 47; Gaps 14;
 31 LATSICASQALRLISIFLGYPAFLFYHYFYKE--LYLLHLEFTFTGIAVFNFN 88
 22 LADWGLSLVDVNFILICQISALFLASLFSMLHPSVSKLHRTALSLGLAFHCEQ 81
 89 QLYH----SLLCIVLQFLILRLMGRTTAVLTTF--CFOMAYLLAGYYTATGNDIKW 141
 82 QAIHLAGLPAICIVIRTPQPRIVQPAVLVAMSYYLLCVHMRQLYDY-----GSYALDI 136
 142 TMRHCVLTKLIGLVANDYFDG--GKQNSLSSQOKAIVGSLVLEVAFSFYGAFLVG 200
 137 TGPLMITTQVKSVLASFHDFVRGDELTKAQOAHAIKMPSSALEFSGYVHNPSTILAG 196
 201 POFSMNHMYKLVQ--ELIDIP---GKIPNS-----IIP-----ALKRLSG-----LFYLVG 243
 197 PLVFKDYIEFVPEGNYLSTPPGNGVLDSSKREVLIEPPTAIVRKVGSVLCAFIPMK 256
 244 VTLTSP--HITEDYLLTEDYDHNHPFRCMYLIMGKFLVLYKVTCTWLTGVCILTGIG 301
 257 FKIVYVVKMKEDDFFNNNTSMYKYV---YAMMATTCIRFKYAHMLADAI CNNSGIG 312
 302 FNGFEKGAQKADACANMKVLFETNPRPTGTITASNITNMAVARIKRLKFLGNKL 361
 313 FPGYDGDGSKWDLISNINVLSEFESTNRDAMNNNCGTNRRLRTLVYERP---QQY 368
 362 SQGLSLFLALMHGSLGSLVCFQMFELVIVEROARLIQSPSTLSKLAATVLOPFYX 421
 369 GILTLTALAAVHMGFPFGYLLTFAATGAVVTAARTKRLFRH-----RQOSTVTRMFYD 423
 422 LVQQTILHLMFGVSMATFCLFTWDMKLKYKSIYFLGHIFFLSLFLP 470

DB 424 ILTCITRVLVGATFPFVLLEFGMSIKLYRFLYLCIHTISLVTIFLP 472
 RESULT 10
 ID 09CAN8 PRELIMINARY; PRT; 465 AA.
 AC 09CAN8;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN F16M19.14 OR AT1G63050.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eustoidia II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ccv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Etmu P., Feldjany T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldenrich A.D., Haas B., Hansen N.P., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marzilli A.,
 RA Miltseher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Southwick A.M.,
 RA Sakano H., Salzenig S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D.,
 RA Ueberck T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
 RT "Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*.";
 RT Nature 408:816-820(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Omdera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carrinci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Meyers A.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
 RA Davis R.W., Ecker J.R., Theologis A.,
 RT "Full length cDNA of gene At1g63050 (GI:15221667).";
 RT Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Deng J.M., Hsuan V.W.,
 RA Lee J.M., Quach H.L., Tang C., Toriumi M., Wu H.C., Yu G., Bowser L.,
 RA Carrinci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.,
 RT "Arabidopsis Open Reading Frame (ORF) Clones.";
 RT Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC010795; AA051612.1;
 DR EMBL: AY072080; AA059903.1;
 DR EMBL: AY122979; AA067512.1;
 DR InterPro: IPR001064; Cysteal1in.
 DR InterPro: IPR004299; MBOAT_fam.
 DR Pfam: PF03062; MBOAT; 1.
 DR PROSITE: PS00225; CRYSTALLIN_BETACAMMA; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 465 AA; 52769 MW; 9A35E91B8277DCB3 CRC64;

Query Match 15.8%; Score 407; DB 10; Length 465;
 Best Local Similarity 25.4%; Pred. No. 2.6e-26;
 Matches 123; Conservative 90; Mismatches 227; Indels 44; Gaps 12;

QY 26 LSLKATLSLGASEOALRLIISIFLGYPALFYRYHYLYKETYLIHLPHFTGTSLAYFN 85
 DB 4 LDMMSAASIGVSVLRFLLCFVATIPISFLMR--FLPSRLGHIYSAASGALSYLS 60
 QY 86 FGNOLYHSLCTIVLQFLRLMGRTITAV-----LTFPCQMAVLLAGYTTATGN-- 137
 DB 61 FG-----FSSNLFPLVPMPTIGVASMAYRPLSGFTTFELGFAVLICGHVFMGDAWK 113
 QY 138 --DIKMPHCVTLTKLIGLAVDYFDGKONSLSSEQKYAIRGVSLLLEVAGSFYFG 195
 DB 114 EGGIDSTALMVLTKVISCINVDGMLKEGLAEAKKRLIQLMPSLIEFGCLCCG 173
 QY 196 AFLVGPQSMNHMYKLVOGE--LIDIPGKIPNSIIPALKRISLGLFYLVGTYLLSPHT 252
 DB 174 SHFAPRVPEMKDYLETWTEKGIWAVSEKGRPSPYGAMIRAVFOALCMALYLVPOFP 233
 QY 253 EDYLLTEDYDNHPEFRCMYMLIWGKFVLYKYVTGMLVTEGVCILTGFGNPF--EEKGK 310
 DB 234 LTRFTEPYQEKGFELKRFQYQWAGFTYARKYFTIWSISEASTIISGLFGSMDTETQK 293
 QY 311 AKMDACAMKWLFEFTNPREFTGTIASFNINTNAMYARYIFKRLKFLGNKE-----LSQG 364
 DB 294 AKMDAKXVDILGLVLAASAQVIRPLEWNIQVSTWLRHYERYIVRGKAGFOLLATQT 353
 QY 365 LSLFLALHGLHSGYLVCQFQMEFLIYVERQARLIQ--SEPTLSKLAITVLOPFYLYV 423
 DB 354 VS---AVMHGLYPCYIIFEVQALMIDGSKAIYRMOQAIPPKAMLRNVLLVLEFLTV 409
 QY 424 QOTIHWLMGYSMTAFCLFTMDKMLKYKSYIFLGHIFLSLFLPIPHKAMVRKEXL 483
 DB 410 -----VVLNYSVGFVWLSLHETLVAFKSVYIGTIVPIAVL--LLSYL--VPVKVRPKT 461

QY 484 KME 487
 DB 462 RKE 465

RESULT 11
 Q42916 PRELIMINARY; PRT; 509 AA.

AC 042916;
 DT 01-JAN-1999 (TREMBLrel. 09, Created)
 DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN SPBC16A3.10.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972h;
 RA Wood V., Rajandream M.A., Barrell B.G., Beck A., Reinhardt R.
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL021748; CAA16861.1; -
 DR GeneDB; Spombe; SPBC16A3.10; -
 DR InterPro; IPR004299; MBOAT_fam.
 DR Pfam; PF03062; MBOAT; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 509 AA; 58821 MW; 74B4E3E3D9DE65D9 CRC64;

Query Match 15.7%; Score 406; DB 3; Length 509;
 Best Local Similarity 25.1%; Pred. No. 3.5e-26;
 Matches 121; Conservative 86; Mismatches 206; Indels 70; Gaps 13;

QY 25 ELINKATLSLGASEOALRLIISIFLGYPALFYRHY-----LFYKETYLI--HLPHFTGTSLI 69

DB 6 DIPEYFSSFLGVHPDQLKLLFCFLSAVFFGILKRLPSAPWIRNLFSSISGLF-----YL 61
 QY 70 IHLPHFTGTSLIAFN-----FGNOLYHSLCTIVLQFLRLMGRTITAVLTTFCQWAX 124
 DB 62 IGVAHLDGVLLVLFDDLFTYFAAFRRSSMPWIIFFV--ILGHTSSHVIRYIY---- 115
 QY 125 LLAGYVYATATNVDYIKMTMPHCVTLTKLIGLAVDYFDGKONSLSSEQKYAIRGVSPL 184
 DB 116 -----PSNTIPI--TASQWULCMKLTARMSVYDRLSSLSSTYQKRALRIPINI 165
 QY 185 LEVAGFSYFCAFLVGPQSMNHMYKLVO---GELIDIPEK--IPNSIIPALKRISLGL 238
 DB 166 LYFLGYVFFPSLLVGPFAFDVDERFTLSMFKPLADPYEKQTPHSLBPALGRCMRGL 225
 QY 239 FYLVGTYLLSHIREDYLLTEDYDNHPEFRCMYMLIWGKFVLYKYVTGMLVTEGVCILT 298
 DB 226 LMLLFTGTSSIVPLKFLTPKFASSPILLKYGVYCTAFAARMKYGAMELSDGACILS 285
 QY 299 GLGFNGFEKAKMDACAMKWLFEFTNPREFTGTIASFNINTNAMYARYIFKRLKFLGN 358
 DB 286 GIGINGLDSGHPMDRYKNIIDPIKEFADNIKALEAMNNTKMLRNYIYLAARAKGK 345
 QY 359 KE-LSQSLFLFL-ALMHGLHSGYLVCQFQMEFLIYVERQARLIQ-----ESPTLS 408
 DB 346 RPFKSTLSTFTVSAAMHGVAGYLLFVSAFIQTVAKYTRRHVRPFPLKPMDETPE--- 402
 QY 409 KLAITVLOPF---YLVQOTIHWLMFGYSMTAFCLFTMDKMLKYKSYIFLGHIFLSL 465
 DB 403 -----GPKRYVDYIGVATNLSLSTYLIISFLNLKESIHVMKELYPIVHIYILIA 454
 QY 466 LFI 468
 DB 455 LAV 457

RESULT 12
 Q08548 PRELIMINARY; PRT; 619 AA.

AC 008548;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Chromosome XV reading frame ORF YOR175C.
 GN YOR175C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE OF 269-619 FROM N.A.
 RA Bordonne R., Camasses A., Madania A., Martin R.P., Poch O.
 RA Tarasov I.A., Winsor B.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Hughes B., Pohl T.M.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC MIRS;
 RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; 275083; CAA99384.1; -
 DR SGD; S0005701; YOR175C.
 DR InterPro; IPR004299; MBOAT_fam.
 DR Pfam; PF03062; MBOAT; 1.
 SQ SEQUENCE 619 AA; 72227 MW; 87C7AC9194BE0BC5 CRC64;

Query Match 15.7%; Score 406; DB 3; Length 619;
 Best Local Similarity 26.2%; Pred. No. 4.5e-26;
 Matches 133; Conservative 79; Mismatches 238; Indels 58; Gaps 17;

QY 28 LNKATLSLGASEOALRLIISIFLGYPALFYRHY-----LFYKETYLI--HLPHFTGTSLI 81

Db 9 LKRLITNYGIDSEFTLRYAICLGSFPLNALIKRIPEKRIGLKCCFIISMSEFYL----- 63
 QY 82 AAFNFGNOYHSLCLCIVLOFLRLMGRITTAULTFFCQOMAYLAGYTTATGNDIK 141
 Db 64 GVALNVSGRRLTPISTMTFYLLISRFYRSKMPHL-NFMVVGHLAINHIAOFLMEQT 122
 QY 142 TM-----PHCVTLTKLGLAVDYDGG-----KDQNSISEQOKAIRGVSLLEVAGFSYF 193
 Db 123 TYDISSQNVLMAMKLTSFAMSYDGSCTESDPKDLTEHOKRAVGHGHPPLKFLAYAF 182
 QY 194 YAAFLVGPQFSNMHYKLVQGEI-DIP-----GKIPNSIIPALKRSLGL 238
 Db 183 YSTLLTGPSPDYADFDMSINCEMFRLPESSKKPMRRHHGERRQIPKNGKALMKVYQGL 242
 QY 239 FVLVGTLLSPHITTELYLTED-YDNHPWFRCMYMLNGKFLYKYVTCWLTBEVCVL 297
 Db 243 AMWILSTLGMKHPVRYVLDKQGFPRSEIFRIHYLLFGFIHRFYAAMTISEGSCYL 302
 QY 298 TGLGFGFEEK-GKAKMDACAMKVMLEFNPRFTGSIASFNINNAVARYIFKRLKFL 356
 Db 303 CGLGVNGYDSTOKIRMDVRNIDITVETAQNTREMLAMNNTKMLKYSYILAVTK 362
 QY 357 GKKEISQGLSLFL-ALMHGLSGYLVCFQMEFLIVEROAARLIQES--PTSKLAA 412
 Db 363 GKPGFRSTLFTFLTSAPMHGTRPGYLLTFATGALT-----QTCGKIYRNNRPIFLBEDG 418
 QY 413 ITVL--QPFYLVQOQTHLMFGYSMTAFCLFTWDMKLVKYSIYLGHIIF--FLSLFI 468
 Db 419 VPLPSKTIYDLGVIYALAFGYMVOPIILDLKPSLWWSVYVYHIVAFSPFLR 478
 QY 469 LPY-----IHKAMVPRK--EKLKME 487
 Db 479 GYAKQVTEPFPSKQPKEIFIRKQKLE 506

RESULT 13

Q8RMH4 PRELIMINARY; PRT; 462 AA.
 AC 08RMH4
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Hypothetical protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 GN AT1G12640.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxId=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Bower L., Jones T., Bach J., Carlinici P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shin P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Tripp M., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,
 RA Palm C.J., Bower L., Jones T., Bach J., Carlinici P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shin P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY093087; AAM913086.1; -
 DR EMBL; AY128762; AAM91162.1; -
 DR InterPro; IPR004299; MBOAT_fam.
 DR Pfam; PF03062; MBOAT; 1.
 KW Hypothetical protein.

SQ SEQUENCE 462 AA; 52151 MW; 3817A68E0CB35A5 CRC64;

Query Match 14.9%; Score 383.5; DB 10; Length 462;

Best Local Similarity 25.6%; Pred. No. 2.6e-24;

Matches 125; Conservative 86; Mismatches 215; Indels 59; Gaps 13;

QY 26 LSLNLTATSLGSEQALRLIISIFLGYPFALFYRHLYPEKETYLLIHLPHFTGSLAYFN 85
 Db 1 MDMSMASIGSVAVLHFLCFVATIVPSACR---IVPSRLGHLVLAASGAVLSLUS 57
 QY 86 FGNQYHSLCLCIVLOFLRLMGRITTAULTFFCQOMAYLAGYTTATGNY-- 137
 Db 58 FG-----FSSNLHFLVPMITGYASMAIYRKCIIITFFGLFALICGVFYMGGDAWK 110
 QY 138 --DIKMTPHCVTLTKLGLAVDYDGGKDQNSISEQOKAIRGVSLLEVAGFSYFYG 195
 Db 111 ECGIDSTALNVLTKLVISCSNNYDGMLEKSGLEAKRNLQIOMPSLIRFGYCLCG 170
 QY 196 AFLVGPQFSNMHYKLVQGEI-----DIPGKIPIPSIIPALKRSLGLFYLGYT 245
 Db 171 SHFAPGVYEMQDYLEWTEGKGIWDTTERKKRPSYCATIRAILQALICMALYLVLPQY 230
 QY 246 LISPHITDYLITEDYDNHPWFRCMYMLNGKFLYKYVTCWLTBEVCILTGLGNF 305
 Db 231 L--TRFTPEPV-----YQEMGFPRKPSYQYMGFTARMKYFIWISSEASIIISGIFSGW 283
 QY 306 BEKG--KAKMDACAMKVMLEFNPRFTGSIASFNINNAVARYIFKRLKFLNKE-- 360
 Db 284 TDDASPKRMDRANKVDILGVELAKSAVOIPLWMIQVSTWLRHYVRLVONGKAGFF 343
 QY 361 --LSQGLSLFLALMHGLSGYLVCFQMEFLIVEROAARLIQES--PTSKLAAITVL 416
 Db 344 QLLATQVTS---AVMHGILPGYMMFVQSAIIMGSRVYIRWQOAIIPKAMLNINWF 399
 QY 417 QPFYLVQOQTHLMFGYSMTAFCLFTWDMKLVKYSIYLGHIIFSL--FLIPYH 473
 Db 400 INFLYTV-----LVLNYSVAGFVWLSLHETLTAYGSVYIGTIIIPGLIILSVVPAKP 453
 QY 474 KAMVPRKEX 482
 Db 454 SRPKRKEE 462

RESULT 14

Q18856 PRELIMINARY; PRT; 480 AA.
 AC Q18856
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE C54G7.2 protein.
 GN C54G7.2
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoa; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodetinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berka M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kerhaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McKerrill A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownteen R.,
 RA Smaiden N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterson R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Mohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 RN [2]

RP SEQUENCE FROM N.A.
 RA Du 2.1;
 RT "The sequence of C. elegans cosmid C54G7."
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U40410; AAA81391.1; -
 DR WormRep; C54G7.2; CE04270.
 DR InterPro; IPR004299; MBOAT_fam.
 DR Pfam; Pf03062; MBOAT; 1.
 DR SQUENCE 480 AA; 55273 MW; 2229DA04ADDF5B55 CRC64;
 Query Match 14.9%; Score 383.5; DB 5; Length 480;
 Best Local Similarity 24.1%; Pred. No. 2.7e-24;
 Matches 112; Conservative 98; Mismatches 221; Indels 33; Gaps 13;
 QY 28 LNKLTATSLGSEQLRLIISIFLGPFPALFYRHYLFYKERT--YLHLPHFTPTGLSIAYFN 85
 DB 14 LEPLAETIGIEADRVNVLISLFCFGISVYRKTCSLKQVNRQVTVPAVVGIVGIVFC 73
 QY 86 FGNQVHSLCTIVLQFLILMGRT-ITAVLTFFCFQMAVLLAGYYTTATGNDIKWTMP 144
 DB 74 FGRALKILANALGSYAIMYFAPPTOVHKTIVLFSMGVLPFHICYRMALLESYSLDVTGP 133
 QY 145 HCVLTLLKLGAVDPFG-GKDONSLSSBOOKYAIRGVPSLLEVAGSFYGAFLVGPQF 203
 DB 134 IMVAVEKATMMAFNLKDGAKADSKLTBEQRESLKETPSLSEMSFMENQIVLTGPAN 193
 QY 204 SMNHMYMLVOGE--LIDIPKIPNSIIPALKRSLGLFYLV--VGYTLLSPHITBEDYLTED 260
 DB 194 NYVDYILFDEBKHLVADKHGLSPPTGAAMKKFYQSIFLAIIVTLGGKTKVED-VGTGP 252
 QY 261 YDNHFF--WRRCMYMLMGKFLVLYKTYTCMLVTEGVCILTLGNGEPEKGAADKACAN 318
 DB 253 YFALPFPQWPFWMFTITF--FIRCAVYFAMVFADAIQMSGFCGSGYDKGNEMKCLTN 310
 QY 319 MKWALFPTNRFPTGIIASFINNTAMVARYIFKELKLGKNEKLSQGLSLFLMLHGLHS 378
 DB 311 VLPEVQEMASQKETLGGMINIQTFWLRKGYER---APKSITVATVYTLISAVHGVSI 366
 QY 379 GYLVCFQMEFLIVVERQARLIOES-----PTLSKLAITVLOPFYVLTQTHLMFM 432
 DB 367 GYVWAFPTGIFLY---AAQTFRSMRWKRFDDHKKFADI---FSFIISK---IAL 415
 QY 433 GYSMTAFCLFTWDKMLKYKSIYFLGHIFFLSLFIPIYHKAM 476
 DB 416 AYATYSFVTWMMYPAVDVLKRVYFPIHIVAFVIFALPKFFKPL 459
 RESULT 15
 Q9LN83 PRELIMINARY; PRT; 457 AA.
 AC Q9LN83;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE T12C24.17;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chao Q.; Brooks S.; Buehler B.; Johnson-Hopson C.; Khan S.; Kim C.;
 RA Shum P.; Altafi H.; Bei Q.; Chin C.; Chio J.; Choi E.; Conn L.;
 RA Conway A.; Gonzales A.; Hansen N.; Howng B.; Koo T.; Lam B.; Lee J.;
 RA Lenz C.; Li J.; Liu A.; Liu K.; Liu S.; Mukharsky N.; Nguyen M.;
 RA Palm C.; Pham P.; Sakano H.; Schwartz J.; Southwick A.; Thavert A.;
 RA Toriumi M.; Vayberg M.; Yu G.; Federspiel N.A.; Theologis A.;
 RA Ecker J.R.;

RT "Genomic sequence for Arabidopsis thaliana BAC T12C24 from chromosome
 I."
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC025417; AAF88094.1; -
 DR InterPro; IPR004299; MBOAT_fam.
 DR Pfam; Pf03062; MBOAT; 1.
 DR SQUENCE 457 AA; 51600 MW; BB4E4BF4F2A4BD CRC64;
 Query Match 14.8%; Score 382.5; DB 10; Length 457;
 Best Local Similarity 26.0%; Pred. No. 3.1e-24;
 Matches 126; Conservative 82; Mismatches 217; Indels 59; Gaps 13;
 QY 31 LATSLGSEQLRLIISIFLGPFPALFYRHYLFYKERT--YLHLPHFTPTGLSIAYFNQNL 90
 DB 1 MASGISVAVLFLCFAVATIVSFACR---IVPSRLGHLVLAASGATLSLFSG--- 54
 QY 91 YHSLCTIVLQFLILMGRTIV---LTPFCFQMAVLLAGYYTTATGNY---DIK 140
 DB 55 ---FSSNLHFLVPMITGYASMAIYRPKCGIITFFLGFAYLIGCHVFYMSGDMKEGID 110
 QY 141 WTPHCVLTLLKLGAVDPFGGKDONSLSSBOOKYAIRGVPSLLEVAGSFYGAFLV 200
 DB 111 STALMVLTLKVLISCSMNINDGMLKEGLREAOQRKRLIOMPSLIEYFGCLCGSHFAG 170
 QY 201 PFSNMHMYMLVOGELI-----DIPKIPNSIIPALKRSLGLFYLVGTYLLSPH 250
 DB 171 PVYEMDYLEMTGKGIWOTTEKRRKPSRYGATIRALQALICMALVLYVQYPL--TR 228
 QY 251 ITEDYLTEDYDNHPPFRCMYMLMGKFLVLYKTYTCMLVTEGVCILTLGNGEPEKGA 309
 DB 229 FTEPV---YQWGFPRKPSYQYMGFTARMXYFPIWSISEASIIISGLGFSGWTGDAS 283
 QY 310 -KXKMDACAMKWLPEPTNRFPTGIIASFINNTAMVARYIFKELKLGKNE-----LS 362
 DB 284 PKRMDAKKVDLIGVELAKSAVOIPLWNIQVSTWRHYVEELVONGKAGPFOLLAT 343
 QY 363 QGSLFLALMHLGSGYLVCFQMEFLIVVERQARLIOE--SPTLSKLAITVLOPFY 421
 DB 344 QIVS---AVMHGLYFGYMMFVQSALMINGSVITYWQQAISPKYMLNINIFINFLY 399
 QY 422 LVQTHLMFMGYSMTAFCLFTWDKMLKYKSIYFLGHIFFLSL---FIPIYHKAMVP 478
 DB 400 TV-----LVNYSAVGFMVLSHETLTAYGSYYIGTIIIPVGLILSYVPAKSRPKP 453
 QY 479 RKEX 482
 DB 454 RKEE 457

Search completed: November 21, 2003, 14:16:04
 Job time : 41 secs

THIS PAGE BLANK (USPTO)

PN WO200008155-A2.
 XX 17-FEB-2000.
 XX
 XX
 PF 06-AUG-1999; 99WO-US17777.
 XX
 XX 07-AUG-1998; 98US-0160065.
 PR 01-SEP-1998; 98US-0098703.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Hillman JL, Yue H, Lai P, Tang YT, Gorgone GA, Guegler KJ;
 PI Corley NC, Baughn MR;
 DR MPI: 2000-205710/18.
 DR N-PSDB; AAF250889.
 XX
 PT New human receptor-associated proteins (HRAP) useful for the diagnosis,
 PT treatment and prevention of cell proliferative, autoimmune,
 PT inflammatory, reproductive, cardiovascular, and gastrointestinal
 PT disorders -
 XX
 PS Claim 1; Pages 73-74; 99pp; English.
 XX
 CC The present sequence is a human receptor-associated protein
 CC (HRAP) from incyte clone 2906971 obtained from THYMOT05 cDNA library.
 CC This sequence is expressed in nervous, gastrointestinal
 CC and reproductive tissues. HRAP has cytostatic, immunomodulatory,
 CC antiinflammatory, cardiac, antihypertensive, hepatotropic,
 CC antitumorigenic, antineuritic, osteopathic, antiallergic, antianemic,
 CC antidiabetic, antidiabetic, dermatological and neuroprotective
 CC activities. The present sequence is useful in the diagnosis, treatment
 CC and prevention of disorders associated with HRAP expression, especially
 CC cell proliferative, autoimmune/inflammatory, reproductive,
 CC cardiovascular and gastrointestinal disorders (e.g. atherosclerosis,
 CC cirrhosis, leukaemia, cancer, AIDS, arthritis, allergies, anaemia,
 CC asthma, dermatitis, diabetes, osteoporosis, multiple sclerosis and
 CC irritable bowel syndrome).
 XX
 SQ Sequence 487 AA;
 Query Match 100.0%; Score 2580; DB 21; Length 487;
 Best Local Similarity 100.0%; Pred. No. 5.5e-275;
 Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 MASSAGDEGTVALAGVQSGFQELSLNKATSLGASEQALRIISIFLGYPPALFYRH 60
 QY 1 MASSAGDEGTVALAGVQSGFQELSLNKATSLGASEQALRIISIFLGYPPALFYRH 60
 Db 61 YLFYKETYLIHLPHFTGTSLIAFNGNQLYHSLGICIVQFLIRLMGRITFAVLTTCF 120
 QY 61 YLFYKETYLIHLPHFTGTSLIAFNGNQLYHSLGICIVQFLIRLMGRITFAVLTTCF 120
 Db 61 YLFYKETYLIHLPHFTGTSLIAFNGNQLYHSLGICIVQFLIRLMGRITFAVLTTCF 120
 QY 121 QMAVLLAGYYTATGNYDIKMTWPHCVLTLLKLGAVDYFDGKQNSLSSSEQKXAIRG 180
 Db 121 QMAVLLAGYYTATGNYDIKMTWPHCVLTLLKLGAVDYFDGKQNSLSSSEQKXAIRG 180
 QY 121 QMAVLLAGYYTATGNYDIKMTWPHCVLTLLKLGAVDYFDGKQNSLSSSEQKXAIRG 180
 Db 121 QMAVLLAGYYTATGNYDIKMTWPHCVLTLLKLGAVDYFDGKQNSLSSSEQKXAIRG 180
 QY 181 VQSLLEVAGSYVYGAFVVGPOFSNMHMYKLVOGELIDIPGKIPNSIIPALKSLGLRY 240
 Db 181 VQSLLEVAGSYVYGAFVVGPOFSNMHMYKLVOGELIDIPGKIPNSIIPALKSLGLRY 240
 QY 181 VQSLLEVAGSYVYGAFVVGPOFSNMHMYKLVOGELIDIPGKIPNSIIPALKSLGLRY 240
 Db 181 VQSLLEVAGSYVYGAFVVGPOFSNMHMYKLVOGELIDIPGKIPNSIIPALKSLGLRY 240
 QY 241 LVGYTLLSPHITEDYLTEDYDNHPRFPCMYMLINGKFLYYKVCMLTEGVCILTGL 300
 Db 241 LVGYTLLSPHITEDYLTEDYDNHPRFPCMYMLINGKFLYYKVCMLTEGVCILTGL 300
 QY 241 LVGYTLLSPHITEDYLTEDYDNHPRFPCMYMLINGKFLYYKVCMLTEGVCILTGL 300
 Db 241 LVGYTLLSPHITEDYLTEDYDNHPRFPCMYMLINGKFLYYKVCMLTEGVCILTGL 300
 QY 301 GFNGFEKGAKKADACANNKVMLEFETNPRPTGTIASFNINTANWARYIFKRLKFLGNKE 360
 Db 301 GFNGFEKGAKKADACANNKVMLEFETNPRPTGTIASFNINTANWARYIFKRLKFLGNKE 360
 QY 301 GFNGFEKGAKKADACANNKVMLEFETNPRPTGTIASFNINTANWARYIFKRLKFLGNKE 360
 Db 301 GFNGFEKGAKKADACANNKVMLEFETNPRPTGTIASFNINTANWARYIFKRLKFLGNKE 360
 QY 361 LSGGLSLFLALMHGHSGLVCFQMEFLIVYERQARLIQSSPTLSKIAATTVQAPY 420
 Db 361 LSGGLSLFLALMHGHSGLVCFQMEFLIVYERQARLIQSSPTLSKIAATTVQAPY 420

QY 421 YLVQOTIHLMPGYSMTAFCLFTWDMKMLKYKYSIYFLGHIFLFLPYLHKAMVPRK 480
 Db 421 YLVQOTIHLMPGYSMTAFCLFTWDMKMLKYKYSIYFLGHIFLFLPYLHKAMVPRK 480
 QY 481 EKLKQME 487
 Db 481 EKLKQME 487
 RESULT 2
 AAB31669
 ID AAB31669 standard; Protein; 487 AA.
 XX
 AC AAB31669;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Amino acid sequence of a human protein having a hydrophobic domain.
 XX
 KW Human; hydrophobic protein; secretory protein; membrane protein; sepsis;
 KW tumour inhibition; immune deficiency; autoimmune disorder; anaemia; burn;
 KW infectious disease; cancer; ulcer; periodontal disease; coagulation;
 KW Parkinson's disease; fertility; immune response; thrombosis.
 XX
 OS Homo sapiens.
 XX
 PN WO200104297-A2.
 XX
 PD 18-JAN-2001.
 XX
 PF 16-JUN-2000; 2000WO-JP03942.
 XX
 PR 08-JUL-1999; 99JP-0194359.
 XX
 PA (SAGA) SAGAMI CHEM RES CENT.
 PA (PROT-) PROTEGENE INC.
 XX
 PI Kato S, Kimura T;
 PS
 DR MPI: 2001-103081/11.
 DR N-PSDB; AAF25159; AAF25169.
 PT
 PT Isolated human proteins and polynucleotides are used in research and
 PT have activities including cell proliferation/differentiation activity,
 PT immune stimulating activity and receptor/ligand activity -
 XX
 PS Claim 1; Page 90-93; 151pp; English.
 XX
 CC The present sequence represents a human protein with hydrophobic domains.
 CC The protein possesses a hydrophobic domain and so is a secretory protein
 CC or a membrane protein. The protein is used as an antigen to prepare
 CC antibodies. The polynucleotide sequence is useful as a source of probes
 CC for genetic diagnosis. It is also useful for producing the protein
 CC in large quantities and for gene therapy. The eukaryotic cells are used
 CC for detecting the receptors or ligands corresponding to the protein and
 CC for detecting small novel pharmaceuticals. The antibodies are also used
 CC for detection, quantification and purification of the proteins. Both the
 CC protein and polynucleotide may be used in research or as nutritional
 CC sources or supplements. The protein may have cytokine and cell
 CC proliferation/differentiation activity, immune stimulating or suppressing
 CC activity, hematopoietic regulating activity, tissue growth activity,
 CC and thrombolytic activity, chemotactic/chemokinetic activity, hemostatic
 CC activity and tumour inhibition activity. It may therefore be used to
 CC treat immune deficiencies resulting from autoimmune disorders or
 CC infectious diseases, cancer, sepsis, anaemias, burns and ulcers or
 CC periodontal disease, Parkinson's disease, induce fertility, improve
 CC immune response and enhance coagulation or inhibit thrombosis.
 XX
 SQ Sequence 487 AA;
 Query Match 100.0%; Score 2580; DB 22; Length 487;
 Best Local Similarity 100.0%; Pred. No. 5.5e-275;

Matches	487; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
QY	1	MASSAEGDEGTVALAGVLOSQFQELSLNKATLSGASEQARLRLIISIFLGYPALFPRH	60		
DB	1	MASSAEGDEGTVALAGVLOSQFQELSLNKATLSGASEQARLRLIISIFLGYPALFPRH	60		
QY	61	YLFYKETYLIHLHFTFTGLSIAYFNGQVHSLCTIYLOFLIRLMGRITTAVALTTTCF	120		
DB	61	YLFYKETYLIHLHFTFTGLSIAYFNGQVHSLCTIYLOFLIRLMGRITTAVALTTTCF	120		
QY	121	QMAVLLAGYVYATANGYDIKMTMPHCVLTLKLIGLAVDFDGGKQNSLSSEQKYAIRG	180		
DB	121	QMAVLLAGYVYATANGYDIKMTMPHCVLTLKLIGLAVDFDGGKQNSLSSEQKYAIRG	180		
QY	181	VPSLLEVGFSYFYGAFLVGPQFSNMHWKLVQSELIDIPKIPNSIIPALKRSLGIFY	240		
DB	181	VPSLLEVGFSYFYGAFLVGPQFSNMHWKLVQSELIDIPKIPNSIIPALKRSLGIFY	240		
QY	241	LVGYTLSPHITTEDYLLTEDYDNHFPWRCMYMLIWGKFLVYKVTCLVTEGVCILTGL	300		
DB	241	LVGYTLSPHITTEDYLLTEDYDNHFPWRCMYMLIWGKFLVYKVTCLVTEGVCILTGL	300		
QY	301	GFNGFEKQKAKMDACANMKWLFETNPFTGTIASFNINNAVARIYFKRLKFLNGKE	360		
DB	301	GFNGFEKQKAKMDACANMKWLFETNPFTGTIASFNINNAVARIYFKRLKFLNGKE	360		
QY	361	LSQGSLLFLALMHLGSHGVLCFQMEFLIYVERQARLQESPTLSKLAITVLOPFY	420		
DB	361	LSQGSLLFLALMHLGSHGVLCFQMEFLIYVERQARLQESPTLSKLAITVLOPFY	420		
QY	421	YLVQOTIHLFMFGYSMTAFCLFTWDKWLKVYKSIYFLGHIFLSLFLPIYHKAMVPRK	480		
DB	421	YLVQOTIHLFMFGYSMTAFCLFTWDKWLKVYKSIYFLGHIFLSLFLPIYHKAMVPRK	480		
QY	481	EKLKQME 487			
DB	481	EKLKQME 487			

RESULT 3

AAB36584 standard; Protein; 487 AA.

QY	1	MASSAEGDEGTVALAGVLOSQFQELSLNKATLSGASEQARLRLIISIFLGYPALFPRH	60
DB	1	MASSAEGDEGTVALAGVLOSQFQELSLNKATLSGASEQARLRLIISIFLGYPALFPRH	60
QY	61	YLFYKETYLIHLHFTFTGLSIAYFNGQVHSLCTIYLOFLIRLMGRITTAVALTTTCF	120
DB	61	YLFYKETYLIHLHFTFTGLSIAYFNGQVHSLCTIYLOFLIRLMGRITTAVALTTTCF	120
QY	121	QMAVLLAGYVYATANGYDIKMTMPHCVLTLKLIGLAVDFDGGKQNSLSSEQKYAIRG	180
DB	121	QMAVLLAGYVYATANGYDIKMTMPHCVLTLKLIGLAVDFDGGKQNSLSSEQKYAIRG	180
QY	181	VPSLLEVGFSYFYGAFLVGPQFSNMHWKLVQSELIDIPKIPNSIIPALKRSLGIFY	240
DB	181	VPSLLEVGFSYFYGAFLVGPQFSNMHWKLVQSELIDIPKIPNSIIPALKRSLGIFY	240
QY	241	LVGYTLSPHITTEDYLLTEDYDNHFPWRCMYMLIWGKFLVYKVTCLVTEGVCILTGL	300
DB	241	LVGYTLSPHITTEDYLLTEDYDNHFPWRCMYMLIWGKFLVYKVTCLVTEGVCILTGL	300
QY	301	GFNGFEKQKAKMDACANMKWLFETNPFTGTIASFNINNAVARIYFKRLKFLNGKE	360
DB	301	GFNGFEKQKAKMDACANMKWLFETNPFTGTIASFNINNAVARIYFKRLKFLNGKE	360
QY	361	LSQGSLLFLALMHLGSHGVLCFQMEFLIYVERQARLQESPTLSKLAITVLOPFY	420
DB	361	LSQGSLLFLALMHLGSHGVLCFQMEFLIYVERQARLQESPTLSKLAITVLOPFY	420
QY	421	YLVQOTIHLFMFGYSMTAFCLFTWDKWLKVYKSIYFLGHIFLSLFLPIYHKAMVPRK	480
DB	421	YLVQOTIHLFMFGYSMTAFCLFTWDKWLKVYKSIYFLGHIFLSLFLPIYHKAMVPRK	480

(INCY-) INCYTE GENOMICS INC.

Yue H, Tang YT, Lal P, Reddy R, Batta S, Baughn MR, Yang J;
Azimzal Y, Lu DM, Au-Young J, Smith LL;WPI: 2001-016234/02.
DR N-PSDB; AAC88075.Human FLEXHT protein and DNA sequences, useful for treating
immunological disorders, developmental disorders, and cancers -
Claim 1; Page 101-102; 166pp; English.

AAC88070 to AAC88124 encode the 55 FLEXHT (full-length molecules expressed in human tissues) proteins given in AAB36579 to AAB36633. The present invention describes an isolated polypeptide (A) comprising an amino acid sequence selected from one of 55 amino acid sequences 42-876 residues in length, corresponding to FLEXHT-1 to FLEXHT-55, a 90 % identical sequence, and a biologically active or immunogenic fragment of the sequence. The FLEXHT proteins can have antianemic, anticonvulsant, antiarteriosclerotic, immunomodulatory, cyrostatic, antistatic, antiinflammatory, hepatotropic, antidiabetic, anti-gout, antihypertoid, neuroprotective, antiarthritic, osteopathic, antipsoriatic, antileuker and antirheumatic activities, and can be used in gene therapy. The polynucleotide sequences can be used to express the protein sequences. Pharmaceutical compositions comprising FLEXHT can be used to treat diseases or conditions associated with altered expression of functional FLEXHT. The proteins and polynucleotides can be used to diagnose and treat disorders, including anaemia, epilepsy, arteriosclerosis, atherosclerosis, developmental disorder, cancers, and immunological disorders such as asthma, bronchitis, cirrhosis, Crohn's disease, diabetes mellitus, gout, Grave's disease, multiple sclerosis, osteoarthritis, pancreatitis, psoriasis, rheumatoid arthritis, and ulcerative colitis.

SQ Sequence 487 AA;

Query Match 100.0%; Score 2580; DB 22; Length 487;

Best Local Similarity 100.0%; Pred. No. 5.5e-275; Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MASSAEGDEGTVALAGVLOSQFQELSLNKATLSGASEQARLRLIISIFLGYPALFPRH	60
DB	1	MASSAEGDEGTVALAGVLOSQFQELSLNKATLSGASEQARLRLIISIFLGYPALFPRH	60
QY	61	YLFYKETYLIHLHFTFTGLSIAYFNGQVHSLCTIYLOFLIRLMGRITTAVALTTTCF	120
DB	61	YLFYKETYLIHLHFTFTGLSIAYFNGQVHSLCTIYLOFLIRLMGRITTAVALTTTCF	120
QY	121	QMAVLLAGYVYATANGYDIKMTMPHCVLTLKLIGLAVDFDGGKQNSLSSEQKYAIRG	180
DB	121	QMAVLLAGYVYATANGYDIKMTMPHCVLTLKLIGLAVDFDGGKQNSLSSEQKYAIRG	180
QY	181	VPSLLEVGFSYFYGAFLVGPQFSNMHWKLVQSELIDIPKIPNSIIPALKRSLGIFY	240
DB	181	VPSLLEVGFSYFYGAFLVGPQFSNMHWKLVQSELIDIPKIPNSIIPALKRSLGIFY	240
QY	241	LVGYTLSPHITTEDYLLTEDYDNHFPWRCMYMLIWGKFLVYKVTCLVTEGVCILTGL	300
DB	241	LVGYTLSPHITTEDYLLTEDYDNHFPWRCMYMLIWGKFLVYKVTCLVTEGVCILTGL	300
QY	301	GFNGFEKQKAKMDACANMKWLFETNPFTGTIASFNINNAVARIYFKRLKFLNGKE	360
DB	301	GFNGFEKQKAKMDACANMKWLFETNPFTGTIASFNINNAVARIYFKRLKFLNGKE	360
QY	361	LSQGSLLFLALMHLGSHGVLCFQMEFLIYVERQARLQESPTLSKLAITVLOPFY	420
DB	361	LSQGSLLFLALMHLGSHGVLCFQMEFLIYVERQARLQESPTLSKLAITVLOPFY	420
QY	421	YLVQOTIHLFMFGYSMTAFCLFTWDKWLKVYKSIYFLGHIFLSLFLPIYHKAMVPRK	480
DB	421	YLVQOTIHLFMFGYSMTAFCLFTWDKWLKVYKSIYFLGHIFLSLFLPIYHKAMVPRK	480

QY 481 EKLKME 487
|||||
Db 481 EKLKME 487

RESULT 4
ABG06198
ID ABG06198 standard; Protein; 424 AA.
XX
AC ABG06198;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #6189.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW Food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
PN WO200175067-A2.
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Dmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR N-PSDB; AAS70385.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 36557; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical.
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 424 AA;

Query Match 87.8%; Score 2265; DB 22; Length 424;
Best Local Similarity 99.5%; Pred. No. 2.3e-240;
Matches 422; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 64 YKEVYLHLFHTFTGSLIAVFNFGNQLVHSLLCIVLQFLILRLMGRITAVLTTCFQMA 123
|||||

Db 1 YKEVYLHLFHTFTGSLIAVFNFGNQLVHSLLCIVLQFLILRLMGRITAVLTTCFQMA 60
QY 124 YLLAGVYVATANGVNDIKMTMPHCVLTLKLGIAVYFDGKQNSLSBQKXVARGVPS 183
Db 61 YLLAGVYVATANGVNDIKMTMPHCVLTLKLGIAVYFDGKQNSLSBQKXVARGVPS 120
QY 184 LLEVAGFSYFYGAFLVGPQPSNMNHYMKLVQGLIDIPKIPNSIIPALKRSLGLFYLVG 243
Db 121 LLEVAGFSYFYGAFLVGPQPSNMNHYMKLVQGLIDIPKIPNSIIPALKRSLGLFYLVG 180
QY 244 YTLSPHITEDYLLTEDYDNHPEFRCMYMLIWGFVLKYKVTCTLVTEGVCITLGLGN 303
Db 181 YTLSPHITEDYLLTEDYDNHPEFRCMYMLIWGFVLKYKVTCTLVTEGVCITLGLGN 240
QY 304 GFEKGRKAKWDACAMMKWLPETNPRFGTASFININAMARIPRKLKFLGKELSQ 363
Db 241 GFEKGRKAKWDACAMMKWLPETNPRFGTASFININAMARIPRKLKFLGKELSQ 300
QY 364 GLSLFLALMGLHSGYLVCFQMEFLIYVERQARLLIQESFTLSKLAITVLQPFYLV 423
Db 301 GLSLFLALMGLHSGYLVCFQMEFLIYVERQARLLIQESFTLSKLAITVLQPFYLV 360
QY 424 QOTIHWLFWGYSMTAFCLFTWDKWLKVKSIYFLGHIFFLSLFLPYIHKAMVRKEXL 483
Db 361 QOTIHWLFWGYSMTAFCLFTWDKWLKVKSIYFLGHIFFLSLFLPYIHKAMVRKEXL 420
QY 484 KME 487
Db 421 KME 424

RESULT 5
ABR71792
ID ABR71792 standard; Protein; 497 AA.
XX
XX ABR71792;
AC
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 42168.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
PN WO200171042-A2.
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
PI
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656680/75.
DR N-PSDB; ABL15895.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 42168; 21pp + Sequence listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB16175), expressed DNA
 CC sequences (AB16176-AB16175) and the encoded proteins
 CC (AB57737-AB57742).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pct_sequences.

XX Sequence 497 AA;

Query Match 33.8%; Score 871.5; DB 22; Length 497;
 Best Local Similarity 37.2%; Pred. No. 11e-86;
 Matches 167; Conservative 92; Mismatches 165; Indels 25; Gaps 5;

QY 28 LNKATSLGASEQALRLIISIFLGFALFYRHVLFKETYLIHLFTFTGLSIAYNPF 86
 Db 14 MDGIASGVPEALRLTLTLTAGYVALVQKFTSVIADKVNHFAGGAGLCYFN 73
 QY 87 GNOYLHSLCTV-LQFLRLMGRITVLTTCFQMAVLLAGYVYTAQNDIKWMPH 145
 Db 74 GLDYHSLIALITLYFLVLLRKTKQIFLAINFVFMGYLLGLGYFTSSNDYDILMTWPH 133
 QY 146 CVLTLKLGLAVDYDGDQNSLSSEQKVALRGVPSLLVAGSYFYGAFLVGPQPSM 205
 Db 134 CILVLRMIGYGFDTGKSESELSKDKETALKKPSPLLLELAFTSPSGFLVGPQPPF 193
 QY 206 NHYMLVQGLIDIFPKIIPNSIIPALKRLSLGLFYLVGYTLLSPHTEYDLYTEDYDNP 265
 Db 194 RRYAFVDFGRFHNGENVAGV---RRFGAGAFLLVYQGLRKLPSTYLTPEFAQVS 249
 QY 266 FWFRCMYLNGKFLVLYRYTCWLVTEGVCILTGSGFNGFEKQAKADACANMKWMLFE 325
 Db 250 FVKRIYLLGFPAKFSLYKYSICWLLTEGALICIGLYKGEDKNGQPDWSGCSNVYKLLLE 309
 QY 326 TNPRTGTLASFNINTNMAVARYLFRKLKFLNGKELSGSLPLALMHGHSGLVCFQ 385
 Db 310 TGNTEHMYVQSFNVNTQWVGQYIKRLKPLNNRTISGALGFVAVHGHSGYMTFL 369
 QY 386 MEFLIVEROAARL-----IOESPTLSKLAITVLQPFYVVOQTTHMLFMGY 434
 Db 370 MEYVNVSTKQITRYTKVLPQMGHILNNSDIYTL-----LYFTLTKSYNVVYMGW 421
 QY 435 SMTAFCLFTMDKMLKVYSIYFLGHIIFPL 463
 Db 422 CLTAFFVFLKRYERMIYVGAVSYGFTFLV 450

RESULT 6
 AAG00235
 ID AAG00235 standard; Protein: 128 AA.

XX AAG00235;
 AC AAG00235;
 XX 06-OCT-2000 (first entry)
 DT 06-OCT-2000 (first entry)
 DE Human secreted protein, SEQ ID NO: 4316.
 XX Human; 5' EST, expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 XX Homo sapiens.
 OS Homo sapiens.
 XX EP1033401-A2.
 PN EP1033401-A2.
 XX 06-SEP-2000.
 PD 06-SEP-2000.
 XX 21-FEB-2000; 2000EP-0200610.
 PF 21-FEB-2000; 2000EP-0200610.
 XX 26-FEB-1999; 99US-0122487.
 PR 26-FEB-1999; 99US-0122487.
 XX (GIST) GENSET.
 PA (GIST) GENSET.
 XX Dumas Milne Edwards J, Duclet A, Giordano J;
 FI Dumas Milne Edwards J, Duclet A, Giordano J;
 XX

DR WPI; 2000-500381/45.
 DR N-PSDB; AAC00241.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX Claim 13; SEQ ID 4316; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 128 AA;

Query Match 24.7%; Score 638; DB 21; Length 128;
 Best Local Similarity 99.2%; Pred. No. 7.7e-62;
 Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MASSAEDDEGTVALACVLOGSPQELSLNKATSLGASEQALRLIISIFLGFALFYRH 60
 Db 1 MASSAEDDEGTVALACVLOGSPQELSLNKATSLGASEQALRLIISIFLGFALFYRH 60
 QY 61 YLFKETYLIHLFTFTGLSIAYNFGNOLYHSLCTVLOFLIRLMGRITTAVALTTCF 120
 Db 61 YLFKETYLIHLFTFTGLSIAYNFGNOLYHSLCTVLOFLIRLMGRITTAVALTTCF 120
 QY 121 QMAYLAG 128
 Db 121 QMAYLAG 128

RESULT 7
 ABB66994
 ID ABB66994 standard; Protein: 722 AA.

XX ABB66994;
 AC ABB66994;
 XX 26-MAR-2002 (first entry)
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 27774.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 OS Drosophila melanogaster.
 XX WO200171042-A2.
 PN WO200171042-A2.
 XX 27-SEP-2001.
 PD 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 23-MAR-2000; 2000US-191637P.
 XX 11-JUL-2000; 2000US-0614150.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR WPI; 2001-656860/75.
 XX N-PSDB; ABL11097.
 DR N-PSDB; ABL11097.
 XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
PT interactions -

PS Disclosure; SEQ ID NO 27774; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 722 AA:

Query Match 15.8%; Score 408.5; DB 22; Length 722;

Best Local Similarity 25.8%; Pred. No. 2e-35; Mismatches 212; Indels 47; Gaps 14;

Matches 121; Conservative 89; Mismatches 212; Indels 47; Gaps 14;

QY 31 LATSIGASBOALRLIISIFLGPALFYRHYLFYKE--TYLIHLFHTFTGLSIAYFNGN 88

DB 22 LADWGLSDVADNPLICQISALFLASLFRSMHPKSVSKLHTFALSGLAFGFCFQ 81

QY 89 QLYH----SLICIVLQFLILRLMGRTTAVLTTF--CFOMAYLLAGYYTATGNDIKW 141

DB 82 QAIHTAGLPAICYIVIRTOPRIVQRAVLVAMSYLLCYHLMRQLYDY----GSYALDI 136

QY 142 TMRPHCVLTKLGLAVDFDG--GKDONSLSSEQKAIKRGVSLLEVAGEFYGAFLNG 200

DB 137 TGRMLITIKVSLFSLHDSFVRGDEELTKAQOYHAIKMPKSALEYSYVHFGSILAG 196

QY 201 POFNMNHYKLVQG--ELIDIP--GKIPNS-----IIP-----ALKSLIG--LFLVWG 243

DB 197 PLVFKYDIEFEGVNLSTPPGNGSLDSKREVLSEPTKAVIRKVGSLVCAFIEMK 256

QY 244 YTLSP--HITDYLITEDYDHNHPFRCMYMLNGKFEVLYKYVTCWLTBEGVCLITGIG 301

DB 257 FVKIYFVKCMKEDDFNMNTSMYKYW---YAMMATTCIRFYKHAMLADAINNSGIG 312

QY 302 FNGFEKSKAKKADACANMKVWLPEINPRFTGTLASNTNMAVARIKRLKFLGNKEL 361

DB 313 FTGYDKDGSKMDLISNINVLSPERSTNMRDAIMNNNCGTNRWLTLYVERVP---QQY 368

QY 362 SOGLSLFLALMHLGSLGVLVCFOMFLIVIVERQARLIQESPTLSKLAITVLQPFY 421

DB 369 GLLTLPALSAVHGFYRGYLLFATGAVVVTARTGRILFRH-----RQSTQVTRMFYD 423

QY 422 LVQQTTHMLFMGYSMTAFCLFTWMDKLYKYSIYFLGHIFFLSLFIPL 470

DB 424 ILTCLITRVVLGYATFPFVLLEFMGSIKILYRFLYLCIHTISVITFIPL 472

RESULT 8

AAU72527

ID AAU72527 standard; Protein; 465 AA.

XX AAU72527;

DT 26-FEB-2002 (first entry)

XX Arabidopsis cell cycle protein CCP14.

XX Cell cycle protein; CCP; cell cycle regulation; herbicide;

XX plant growth regulator; plant development; abiotic stress; biotic stress;

XX nutrient deprivation; pathogen attack; crop yield; immunogen.

XX Arabidopsis thaliana.

PN WO200185946-A2.

XX 15-NOV-2001.

XX 14-MAY-2001; 2001WO-IB01307.

XX 12-MAY-2000; 2000US-204045P.

XX (CROP-) CROPPEDSIGN NV.

XX Inze D, Boudolf V, De Veylder L, Acoستا JAT, Magyar Z;

XX WPI; 2002-062249/08.

XX N-PEDB; AAS96317.

PT New cell cycle protein and nucleic acid molecule encoding it useful for

PT regulating cell cycle progression in plants and for identifying

PT modulators which are useful as herbicides or plant growth regulators -

XX Claim 34; Fig 14; 316pp; English.

XX The invention relates to a novel cell cycle protein (CCP) and the

XX polynucleotides encoding them. CCP is useful for identifying a compound

XX which modulates the activity of the polypeptide and which binds to the

XX polypeptide and an anti-CCP antibody is useful for detecting the presence

XX of CCP in a sample. A CCP modulator is useful for modulating the cell

XX cycle or growth of a plant such as Arabidopsis thaliana, rice, wheat,

XX maize, tomato, alfalfa, oilseed rape, soybean, sunflower and canola.

XX CCP nucleic acid and polypeptide molecules are useful as modulating

XX agents in regulating cell cycle progression in plants. CCP is useful to

XX treat disorders characterized by insufficient or excessive production of

XX CCP protein or production of CCP protein forms which have decreased or

XX aberrant activity. Compounds that bind to or modulate the activity

XX of CCP polypeptide are useful as herbicides or plant growth regulators.

XX The polynucleotide is useful for modifying cell fate, plant development,

XX plant morphology, biochemistry and/or physiology, the length of the G1,

XX S, G2 and/or M phase of the cell cycle of a plant, initiation, promotion,

XX stimulation or enhancement of cell division, DNA replication, seed set,

XX seed size, seed development, tuber, fruit, leaf formation, shoot and root

XX initiation and/or development, nodule function, dwarfism in plants,

XX senescence, tolerance or resistance to stress. CCP, the polynucleotide

XX and the anti-CCP antibody are useful in agriculture to modulate the

XX CCP protein levels or activity of a protein involved in the cell cycle due

XX to environmental conditions, including abiotic stress such as

XX cold, nutrient deprivation, heat, drought, salt stress, or biotic

XX stress such as pathogen attack, to modulate e.g. enhance crop yields,

XX and attenuate plant architecture, plant quality traits, plant

Sequence 465 AA:

Query Match 15.8%; Score 407; DB 23; Length 465;

Best Local Similarity 25.4%; Pred. No. 1.5e-35;

Matches 123; Conservative 90; Mismatches 227; Indels 44; Gaps 12;

QY 26 LSLNKLATSLGASBOALRLIISIFLGPALFYRHYLFYKE--TYLIHLFHTFTGLSIAYFN 85

DB 4 LDMNSMAISIGSVAVNLFCLFVATITISFLMR---FIPRLGHIITSAASGALSTLS 60

QY 86 FGNOLYHSLICIVLQFLILRLMGRTTAV-----LTFCFOMAYLLAGYYTATGNY-- 137

DB 61 FG-----FSSNHLFTFPMITIGVSMALYPLSGSFTIFPLGFAYLIGCHVFMVSGDAWK 113

QY 138 --DIKTMPHCVLTKLIGLAVDFDGKDNONSLSSEQKAIKRGVSLLEVAGEFYGAFLNG 195

DB 114 BGGIDSTGALMVLTLKIVISINVDGWLKEEGLEAGKRNRLIQMPSLIEFYGLCCG 173

```

QY 196 APLVGFQFSNMHYMKLVQGE---LIDIPKIPNSIIIPAIKRSLGSLGVLYGVYTLSPHIT 252
DB 174 SHFAGVFEMKDYLETETBEKGIWAVSEKGRPSYGAMIRAFQAIQALYLYLVQPP 233
QY 253 EYLLTEDVDNHPFRCMYMLIMGFVLKYVTCMLVTEGICILGLGFNPF--EEKCK 310
DB 234 LTRFEPVYQEMGFLKRFQYQWAGFTARKYFTIWSISEASIIISGLGSGMTDETQK 293
QY 311 AKMDACANNKWLFTETPRFTGLIASFNINTNANVARYIFKRLKFLGNKE-----LSQG 364
DB 294 AKMDAKANVDILGVELAKSAVOIPLFWMIOVETMLRHYEELIVRPGKAGFOLLATOT 353
QY 365 LSLPLALMHGSLGVLVCFQMEFLIVVERQAARLIQF-SPTLSKLAITVLOPFYLYV 423
DB 354 VS---AWHGLYPGYIIFVQSAIMIDGSKAIVRMOQAIPPKMMMLRVALLINFLVTV 409
QY 424 QOTIMLFNGYSMTAFCLFTMDKMLKVYSIYFLGHIFLSLFLPIYHKMVRPKEL 483
DB 410 -----VVLNYSVSGFMVLSLHETLVAFKSVYIIGTVIPAVL-LLSYL-VVKKVRPKT 461
QY 484 KME 487
DB 462 RKE 465

RESULT 9
ABJ19323
ID ABJ19323 standard; Protein; 520 AA.
XX
AC ABJ19323;
DT 28-MAR-2003 (first entry)
DE NOVX related protein SEQ ID No 12.
XX
XX Antidiabetic; anorectic; vitucide; antibacterial; fungicide; nootropic;
XX prozoacide; neuroprotective; antiparkinsonian; antihypaemic;
XX NOVX-associated disorder; metabolic disorder; diabetes; anorexia;
XX obesity; infectious disease; cancer-associated cachexia; immune disorder;
XX neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
XX haematopoietic disorder; cancer; dyslipidaemia; metabolic disturbance;
XX neurogenesis; cell differentiation; cell proliferation; haematopoiesis;
XX wound healing; angiogenesis; gene therapy; chromosome mapping;
XX tissue typing; preventive medicine; pharmacogenomic; NOVX.
XX
XX OS Unidentified.
XX
XX WO200299062-A2.
XX
XX 12-DEC-2002.
XX
XX 04-JUN-2002; 2002WO-US17559.
XX
XX 04-JUN-2001; 2001US-295607P.
XX 06-JUN-2001; 2001US-296404P.
XX 06-JUN-2001; 2001US-296418P.
XX 07-JUN-2001; 2001US-296575P.
XX 11-JUN-2001; 2001US-297414P.
XX 12-JUN-2001; 2001US-297567P.
XX 14-JUN-2001; 2001US-297573P.
XX 14-JUN-2001; 2001US-298285P.
XX 15-JUN-2001; 2001US-298528P.
XX 15-JUN-2001; 2001US-298556P.
XX 21-JUN-2001; 2001US-299499P.
XX 22-JUN-2001; 2001US-300177P.
XX 28-JUN-2001; 2001US-301530P.
XX 28-JUN-2001; 2001US-301550P.
XX 03-JUL-2001; 2001US-302951P.
XX 12-SEP-2001; 2001US-318771P.
XX 25-SEP-2001; 2001US-324687P.
XX 24-OCT-2001; 2001US-339266P.
XX 16-NOV-2001; 2001US-337524P.
XX 14-DEC-2001; 2001US-341143P.

```

```

PR 21-FEB-2002; 2002US-358643P.
PR 21-FEB-2002; 2002US-359151P.
PR 28-FEB-2002; 2002US-361195P.
PR 05-MAR-2002; 2002US-361964P.
PR 10-APR-2002; 2002US-371346P.
PR 10-APR-2002; 2002US-371523P.
PR 03-JUN-2002; 2002US-0161493.
XX
XX (CUBA-) CUBAGEN CORP.
XX
XX Anderson DW, Zethusen BD, Li L, Zhong M, Casman SJ, Gerlach VL,
XX Shinkels RA, Gorman L, Pena CE, Kekuda R, Patutajan M,
XX Spytek KA, Leite MW, Rastelli L, MacDougall JR, Taupier RJ, Guo X,
XX Miller CE, Shenoy SG, Hiale T, Voas EZ, Boldog FL, Malysanek UM,
XX Padigan M, Ji W, Smithson G, Edinger SR, Millet I, Ellerman K,
XX WPI, 2003-140607/13.
XX N-PSDB; ABT16022.
XX
XX New isolated NOVX polypeptides and polynucleotides, useful for
XX preventing, diagnosing or treating NOVX-associated disorders, e.g.
XX obesity, cancer, Parkinson's disease, infections, immune disorders, or
XX various dyslipidemias.
XX
XX Example 6; Page 101; 461pp; English.
XX
XX The invention relates to an isolated polypeptide comprising any of the 36
XX 86-1370 residue amino acid sequences, given in the specification, a
XX mature form of them, or a sequence that is at least 95 % identical to, or
XX having one or more conservative amino acid substitutions in one of the 36
XX amino acid sequences. The polypeptides, nucleic acid molecules and
XX antibodies of the invention are useful in the manufacture of a medicament
XX for treating a syndrome associated with a human disease, preferably a
XX NOVX-associated disorder. The nucleic acid molecules, polypeptides and
XX antibodies are useful for treating, preventing or diagnosing diseases
XX such as metabolic disorders, diabetes, obesity, infectious diseases
XX (viral, bacterial, fungal, helminthic, and protozoal), anorexia, cancer-
XX associated cachexia, neurodegenerative disorders, Alzheimer's disease,
XX Parkinson's disease, immune disorders, haematopoietic disorders, cancer
XX and various dyslipidaemias, or metabolic disturbances associated with
XX obesity, metabolic X syndrome, and wasting disorders. The nucleic acids
XX and polypeptides may also be used as targets for the identification of
XX small molecules that modulate or inhibit e.g. neurogenesis, cell
XX differentiation, cell proliferation, haematopoiesis, wound healing and
XX angiogenesis, in gene therapy, in generation of antibodies that bind
XX immunospecifically to NOVX substances for use in therapeutic or
XX diagnostic methods. The nucleic acids are further used as hybridisation
XX probes, in chromosome mapping, tissue typing, preventive medicine, and
XX pharmacogenomics. This sequence represents a NOVX related protein of the
XX CC
XX CC
XX SQ Sequence 520 AA;

```

```

Query Match 15.1%; Score 389; DB 24; Length 520;
Best Local Similarity 26.1%; Pred. No. 1.7e-33;
Matches 117; Conservative 86; Mismatches 199; Indels 56; Gaps 15;

```

```

QY 55 ALFIRHYLFYKET--YLHLFHTTGLSIAYFNGQLYHSLCLIVLOFLRLMGRTIT 112
DB 39 AIFRTHSHSKTSFIRHVAVTLGLTALFCGMYALHFLVOSGISYCIIMIIIG--V 95
QY 113 AVLTFPCQMA--YL---LAGYYTATGNDIKIMWHCVLTLLKLGADYDPG--GK 164
DB 96 ENMANVCFVPLAGLTVLCQVRRVITFDYQYSAFSGPMIITOKITSLACIHDMGRK 155
QY 165 DONSLSBQKRYAIRGVPSLLEVAGFSYFGAPLVGPQFSNMHYMKLVQGLIDIP---- 220
DB 156 DE-ELTSGQRLAVRBMPSLLEVLSYNCGNFMGILAGPLCSYKDYITTFEGRSYHTSGE 214
QY 221 -GK-----IPNSTIPLAKRLSLGLFLVGVYTLSPHITEDYLLTEDYDNHPPW-F 268
DB 215 NGKETQYERTPEPSNAV-VQKLVVCGLSLLFLHITCTT-LPVEYNIDEHFOATASWPT 272

```


QY 395 ROARLIGES-----PTLSKLAATVLPQPFYLVQOTIHMLFMGYSMTAFCLFTWDKW 447
 DB 359 --AAAAYNNFRHYFIERSQKL-----FYDVTIWLTVQVAISYTVVPLVLSIKPS 408
 QY 448 LKVKYSIYFLGHIIFLSLFLIP 470
 DB 409 LTFYSSWYICHLIGILVLLLP 431
 RESULT 11
 AAB93974
 ID AAB93974 standard; Protein; 432 AA.
 AC AAM93974;
 XX 13-NOV-2001 (first entry)
 DT
 XX Human stomach cancer expressed polypeptide SEQ ID NO 17.
 DE Human stomach cancer; marker; screening; micro-metastasis;
 KW Human; stomach cancer; marker; screening; micro-metastasis;
 XX peritoneal dissemination.
 OS Homo sapiens.
 XX MO200109317-A1.
 PN 08-FEB-2001.
 PD 28-JUL-2000; 2000WO-JP05063.
 PE 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 18-OCT-1999; 99US-0159590.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 17-FEB-2000; 2000US-0183322.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Ieogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Aburatani H;
 PI Kodama T, Midorikawa Y;
 DR WPI: 2001-570287/64.
 DR N-PSDB: AAI93850.
 XX
 PT New Stomach cancer-associated genes, useful as markers in blood tests
 PT for screening for the early stages of the disease -
 PS Claim 1; Page 81-82; 242pp; Japanese.
 XX
 CC The invention relates to stomach cancer-expressed genes
 CC (AAI93842-AAI93917) and the encoded proteins (AAM93967-AAM94039). The
 CC genes can be used as markers in blood tests for screening for the early
 CC stages of the disease. The proteins and peptides can be used as targets
 CC for screening for compounds to treat the disease. They can also be used
 CC for predicting micro-metastases. The gene can predict peritoneal
 CC dissemination.
 CC
 XX Sequence 432 AA;
 SQ
 Query Match 13.0%; Score 334.5; DB 22; Length 432;
 Best Local Similarity 25.7%; Pred. NO. 1.3e-27;
 Matches 98; Conservative 72; Mismatches 163; Indels 49; Gaps 12;

QY 223 -----IPNSIIIPALKSLGLFYLVGTYLSPHITEDYLLTEDYDNHPF--FCWMLI 275
 DB 134 YERTEPSENTAV-VQKLIVCGLSLFLHITCTT-LPVEYNIDEPFOATASPTKIYIYI 191
 QY 276 WGFPELVYKYVTCWLVTGVCILTGLGNGFEKAKKADCAANKWMLFETNPRTGTIA 335
 DB 192 SLAARPRYVPAWTLADAINNAAGFGRGYDENGCAARWDLISNRIQOIEWSTSFKMPLD 251
 QY 336 SFNINTNAWARYIFPKLKFIGNKELSGLSLFLALMHGHSGLVCFQMEPLIVIER 395
 DB 252 NWNITQALMLKRCYKERTSFSPITQ-----FTLSAIGHGVPGYLLFGLVMTL--- 303
 QY 396 QARLIGES-----PTLSKLAATVLPQPFYLVQOTIHMLFMGYSMTAFCLFTWDKW 448
 DB 304 -AARAKRNFRHYFIERSQKL-----FYDVTIWLTVQVAISYTVVPLVLSIKPSL 354
 QY 449 KVKYSIYFLGHIIFLSLFLIP 470
 DB 355 LTFYSSWYICHLIGILVLLLP 376
 RESULT 12
 AAB93797
 ID AAB93797 standard; Protein; 432 AA.
 AC AAB93797;
 XX 26-JUN-2001 (first entry)
 DT
 XX Human protein sequence SEQ ID NO:13560.
 DE Human protein sequence SEQ ID NO:13560.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.
 XX EP1074617-A2.
 PN 07-FEB-2001.
 PD 28-JUL-2000; 2000EP-0116126.
 PE 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Ieogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS Claim 8; SEQ ID 13560; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the protein encoded by
 CC the full-length cDNA. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH0166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 432 AA;

Query Match 13.0%; Score 334.5; DB 22; Length 432;
 Best Local Similarity 25.7%; Pred. No. 1.3e-27;
 Matches 98; Conservative 72; Mismatches 163; Indels 49; Gaps 12;

QY 118 FCFQNAVY-----LAGYVYATGNYDIKWTMPHCVTLTKLIGLAVDPDG--GKQNSLSS 171
 DB 15 FVFALGYLTVCQVTRVYIFDYGYSDAFSGPMWITOKITSLACEIHDMFRKDE-ELTS 73
 QY 172 EQOKAIVGVPSLLEVAGSFYFGAFLVGPQPSNMHMYKLVOGELIDP-----GK----- 222
 DB 74 SGRLAVRMPSLLEVLISYCNFMGILAPLCSYXDYIFIGRSYHITQSGENGKEETQ 133
 QY 223 -----IPRSIIPALKRLSLGLFYLVGYTLSPHITEDYLTLEDYDNHPFM-FRCMYMLI 275
 DB 134 YRTEPSPRTAV-VQKLIVCGISLFLHITCTT-LPVEYNIDENHOATASWPTKIIYLYI 191
 QY 276 WGFVLYVYVTCWLTVEGVCILTGLFNGFEKKAQKADACANMKWMLFETNPRFTGTIA 335
 DB 192 SLAARPKYFFAMTADAINNAAGFGFRGNDGAAKRWDLISNLRIQIEMSTSFPMFLD 251
 QY 336 SFNINTNMAVAVYIFKRLKFLGNKELSGQLSLFLALMHGLSGYLVCQMFLLVIYER 395
 DB 252 MNNIQTALMLKVKVCERTSFSPTIQ-----TPILSAIMHGVYPGYLTFLTGVLMTL--- 303
 QY 396 QAKRLIOES-----PTLSKLAATVLOPFYVLVQOITIMLPMGYSMTAFCLFTWDKWL 448
 DB 304 -AARARNRNFRHYFTEPQKL-----FYDVTIWTVOVAISTVVPVLLSIKPSL 354
 QY 449 KYVKSIVFLGHIFFLSLFLIP 470
 DB 355 TFYSSWYICHLIGILVLLLP 376

RESULT 13

AAAG81345 standard; Protein; 423 AA.

AAAG81345;

10-SEP-2001 (first entry)

Human AFP protein sequence SEQ ID NO:208.

Human; secreted protein; secretion; bacterial cell; fungal cell;
 eukaryotic cell; fusion protein; maltose binding protein;
 immunoglobulin constant region; polystyrene tag.

Homo sapiens.

WO200129221-A2.

26-APR-2001.

20-OCT-2000; 2000WO-US29052.

20-OCT-1999; 99US-0160712.

(ZYMO) ZYMOGENETICS INC.

XX

PI Conklin DC, Yee DP;
 DR WPI; 2001-300340/31.
 DR N-PSDB; AAH52196.
 PT Isolated polypeptide for directing secretion of proteins of interest
 PT from a host cell including, e.g. bacteria, includes contiguous amino
 PT acid residues of polypeptide with specified amino acids
 PS Claim 1; Page 362-364; 617pp; English.

XX AAH52093 to AAH52303 encode the human secreted proteins given in AAH81242
 CC to AAH81453. The secreted proteins can be used for directing the
 CC secretion of proteins of interest from a host cell including bacteria,
 CC fungal cells, and cultured higher eukaryotic cells. The present invention
 CC also describes fusion proteins, where a secreted protein of the invention
 CC is operably linked via a peptide bond or peptide linker to a second
 CC protein selected from the group consisting of maltose binding protein,
 CC an immunoglobulin constant region, a polystyrene tag and a peptide
 CC given in AAH81453.

XX Sequence 423 AA;

Query Match 12.7%; Score 328.5; DB 22; Length 423;
 Best Local Similarity 25.4%; Pred. No. 5.9e-27;
 Matches 97; Conservative 72; Mismatches 164; Indels 49; Gaps 12;

QY 118 FCFQNAVY-----LAGYVYATGNYDIKWTMPHCVTLTKLIGLAVDPDG--GKQNSLSS 171
 DB 6 FVFALGYLTVCQVTRVYIFDYGYSDAFSGPMWITOKITSLACEIHDMFRKDE-ELTS 64
 QY 172 EQOKAIVGVPSLLEVAGSFYFGAFLVGPQPSNMHMYKLVOGELIDP-----GK----- 222
 DB 65 SGRLAVRMPSLLEVLISYCNFMGILAPLCSYXDYIFIGRSYHITQSGENGKEETQ 124
 QY 223 -----IPRSIIPALKRLSLGLFYLVGYTLSPHITEDYLTLEDYDNHPFM-FRCMYMLI 275
 DB 125 YRTEPSPRTAV-VQKLIVCGISLFLHITCTT-LPVEYNIDENHOATASWPTKIIYLYI 182
 QY 276 WGFVLYVYVTCWLTVEGVCILTGLFNGFEKKAQKADACANMKWMLFETNPRFTGTIA 335
 DB 193 SLAARPKYFFAMTADAINNAAGFGFRGNDGAAKRWDLISNLRIQIEMSTSFPMFLD 242
 QY 336 SFNINTNMAVAVYIFKRLKFLGNKELSGQLSLFLALMHGLSGYLVCQMFLLVIYER 395
 DB 243 MNNIQTALMLKVKVCERTSFSPTIQ-----TPILSAIMHGVYPGYLTFLTGVLMTL--- 294
 QY 396 QAKRLIOES-----PTLSKLAATVLOPFYVLVQOITIMLPMGYSMTAFCLFTWDKWL 448
 DB 295 -AARARNRNFRHYFTEPQKL-----FYDVTIWTVOVAISTVVPVLLSIKPSL 345
 QY 449 KYVKSIVFLGHIFFLSLFLIP 470
 DB 346 TFYSSWYICHLIGILVLLLP 367

RESULT 14

ABB64492 standard; Protein; 489 AA.

ABB64492;

26-MAR-2002 (first entry)

Drosophila melanogaster polypeptide SEQ ID NO 20268.

Drosophila; developmental biology; cell signalling; insecticide;
 pharmaceutical.

Drosophila melanogaster.

WO200171042-A2.

XX

PD 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PMD, Myers EM;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL08595.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Disclosure; SEQ ID NO 20268; 21np + Sequence Listing; English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 489 AA;

Query Match 9.4%; Score 241.5; DB 22; Length 489;
 Bect Local Similarity 23.1%; Pred. No. 2,8e-17;

Matches 107; Conservative 73; Mismatches 195; Indels 89; Gaps 17;

QY 77 TGLSIAYENFGNQLVHSLLCIYLOFLILRLMGRITTAV-----LTFECFQAAVL-----L 126
 DB 36 TGLSLVAVVIVISGL-HSLHC-----FVSLALGTAAVLVHPSKGLVTFVAMGCVLVFRI 90
 QY 127 AGYYTATGANDIKMTMPHCIVTLKLIQLAVDPDGKDONSLSBQOKYAIRGVP----- 182
 DB 91 FDFYFGIRGHTN-----MIQMILTLKVSGLAFKETAAMK-RLQAHDEQKNDGRDVGHQPSP 145
 QY 183 -----SLEVAGSEYFYGAFVVGQFSKMHYMKLVQGLIDIPKINSITP 229
 DB 146 IEITDYDELQSLSAEILHVSFNIGVLTGYYRRTYR-----DYFEMPFTYAPFVE 200
 QY 220 A-LKRLSLGLFVLVGYTLSPHITEDYLTEDYDN-HPEWFCMVMILMGKRVLYKYYTC 287
 DB 201 ALLEKTKAAVFYCALYLATNYMPDLYALSDFFPDRSVRYLLV--VWPTFTFRARIY 258
 QY 288 WLVTGVCILTGLGFRNGFEKQ-----KAKMDA-----CANMKVLMFE 325
 DB 259 TGLTISECCCTWAGGAYPDESDPNNGBGRKRYOHLKRDADKHNYNTFTTYNTVLEVE 318
 QY 336 TNPRTGTITASNINNTAMVAVIIFKRLKELCNKELSOGLSLFLALMHGHSYLVCFQ 385
 DB 319 RCMTRREGKHMNVCCQYWLAVNVY---TLFPSKRYRTGATLLCAAYMHGFRPHYFCIM 375
 QY 386 MEFLIYVERQAAARLIQESPTLSKLAATVLOPFYLVQOTTIMLFMGSMN-----APCL 441
 DB 376 GAFPYVSLEDMMDKLVRSATGTSRRVIDV-----TFWIFKMFATFVLSGEAFL 424
 QY 442 FTWDMKLVKYSIYFGLHIFLSL---FILPYIHKAMVPKREX 482
 DB 425 SSFGNIMWRYSVYHIGYISMAAMTALGTYLSQKRAARRKKR 468

RESULT 15
 ABB97850

ID ABB97850 standard; Protein; 256 AA.
 XX
 AC ABB97850;
 XX
 DT 03-OCT-2002 (first entry)
 XX
 DE Human secretory polypeptide (SPTM) 102.
 XX
 KW Human; secretory protein; secretory polynucleotides; SPTM;
 KW severe combined immunodeficiency; intracellular parasite protection;
 KW fungal parasite; protozoan parasite; cell proliferative disorder; cancer;
 KW immune disorder; AIDS; neurological disorder; Parkinson's disease;
 KW motor neuron disorder; demyelinating disease; multiple sclerosis;
 KW meningitis; abscess; prion diseases; cerebral palsy;
 KW neuroskeletal disorder; peripheral nervous system disorder;
 KW dermatomyositis; polymyositis; myopathy; myasthenia gravis;
 KW mental disorder; Tourette's syndrome.
 XX
 OS Homo sapiens.
 XX
 PN WO200220756-A2.
 XX
 PD 14-MAR-2002.
 XX
 PF 30-AUG-2001; 2001WO-US27297.
 XX
 PR 05-SEP-2000; 2000US-229747P.
 PR 05-SEP-2000; 2000US-229748P.
 PR 05-SEP-2000; 2000US-229749P.
 PR 05-SEP-2000; 2000US-229750P.
 PR 05-SEP-2000; 2000US-229751P.
 PR 05-SEP-2000; 2000US-230016P.
 PR 05-SEP-2000; 2000US-230583P.
 PR 06-SEP-2000; 2000US-230585P.
 PR 06-SEP-2000; 2000US-230514P.
 PR 06-SEP-2000; 2000US-230515P.
 PR 06-SEP-2000; 2000US-230517P.
 PR 06-SEP-2000; 2000US-230518P.
 PR 06-SEP-2000; 2000US-230519P.
 PR 06-SEP-2000; 2000US-230595P.
 PR 06-SEP-2000; 2000US-230596P.
 PR 06-SEP-2000; 2000US-230597P.
 PR 06-SEP-2000; 2000US-230599P.
 PR 06-SEP-2000; 2000US-230610P.
 PR 06-SEP-2000; 2000US-230864P.
 PR 06-SEP-2000; 2000US-230865P.
 PR 06-SEP-2000; 2000US-230988P.
 PR 06-SEP-2000; 2000US-230989P.
 PR 06-SEP-2000; 2000US-230990P.
 PR 07-SEP-2000; 2000US-230897P.
 PR 07-SEP-2000; 2000US-230951P.
 PR 07-SEP-2000; 2000US-231163P.
 PR 07-SEP-2000; 2000US-231832P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
 PI Jones AL, Yu JY, Wright RJ, Glezen D, Liu TF, Yap PE, Dahl CR;
 PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
 PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;
 PI Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;
 XX
 DR WPI; 2002-315658/35.
 DR N-PSDB; ABL99847.
 XX
 PT Polynucleotide sequences encoding human secretory proteins useful for
 PT gene therapy of e.g. genetic deficiency disorders, cancers, and
 PT diseases caused by intracellular parasites -
 XX
 PS Claim 29; Page 430-431; 585bp; English.
 XX

Query Match	9.0%	Score 231	DB 23	Length 256
Best Local Similarity	26.3%	Pred. No. 1.5e-16		
Matches 65, Conservative	43	Mismatches 125	Indels 14	Gaps 4

[illegible]

Search completed: November 21, 2003, 14:14:42
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 14:12:23 ; Search time 21 Seconds
(without alignments)
2230.198 Million cell updates/sec

Title: US-09-938-803-6
Perfect score: 2580
Sequence: 1 MASSAECDEGTVALACVLQ.....ILPYHKAMVPRKXKMKME 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_76: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	749.5	29.1	486 2 T15281	hypothetical prote
2	407	15.8	465 2 G96655	unknown protein, 2
3	406	15.7	509 2 T139542	hypothetical prote
4	406	15.7	619 2 S67067	probable membrane
5	383.5	14.9	480 2 T15839	hypothetical prote
6	297	11.5	540 2 T19097	hypothetical prote
7	244.5	9.5	480 2 T20899	hypothetical prote
8	158.5	6.1	366 2 T27912	hypothetical prote
9	147	5.7	135 2 T05909	membrane protein Y
10	116.5	4.5	477 2 AB2313	hypothetical prote
11	116.5	4.5	736 2 C69451	cationic amino aci
12	115	4.5	395 2 B96610	hypothetical prote
13	114.5	4.4	609 2 S65208	probable membrane
14	114	4.4	614 2 B84949	NADH2 dehydrogenas
15	113.5	4.4	547 2 T13437	NADH2 dehydrogenas
16	111.5	4.3	419 2 E90446	permease [imported
17	111	4.3	392 2 A65086	probable hydrogena
18	110.5	4.3	355 2 T13871	NADH2 dehydrogenas
19	110.5	4.3	355 2 T13872	NADH2 dehydrogenas
20	110.5	4.3	355 2 T13873	NADH2 dehydrogenas
21	110.5	4.3	355 2 T13869	NADH2 dehydrogenas
22	110.5	4.3	355 2 T13870	NADH2 dehydrogenas
23	110	4.2	355 2 S49592	cytochrome b-like
24	109.5	4.2	355 2 T13873	NADH2 dehydrogenas
25	109.5	4.2	355 2 T13911	NADH2 dehydrogenas
26	109.5	4.2	355 2 T13912	NADH2 dehydrogenas
27	109.5	4.2	355 2 T14092	NADH2 dehydrogenas
28	109.5	4.2	355 2 T14092	NADH2 dehydrogenas
29	109	4.2	392 2 H91113	hypothetical prote

30	109	4.2	392 2 H85958	hypothetical prote
31	108.5	4.2	355 2 T13986	NADH2 dehydrogenas
32	107.5	4.2	355 2 T13865	NADH2 dehydrogenas
33	107.5	4.2	355 2 T13909	NADH2 dehydrogenas
34	107.5	4.2	355 2 T13910	NADH2 dehydrogenas
35	107.5	4.2	355 2 T13908	NADH2 dehydrogenas
36	107	4.1	355 2 T11734	NADH2 dehydrogenas
37	107	4.1	392 2 AD0885	probable hydrogena
38	107	4.1	619 2 T11314	NADH2 dehydrogenas
39	107	4.1	724 2 T19601	hypothetical prote
40	106.5	4.1	355 2 T13919	NADH2 dehydrogenas
41	106.5	4.1	355 2 T13987	NADH2 dehydrogenas
42	106.5	4.1	355 2 T13984	NADH2 dehydrogenas
43	106.5	4.1	355 2 T13985	NADH2 dehydrogenas
44	106.5	4.1	537 2 T41654	amino acid permeas
45	106	4.1	355 2 T11739	NADH2 dehydrogenas

ALIGNMENTS

RESULT 1

T15281
hypothetical protein R155.1 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C/Accession: T15281

R/Getset, C.; Mameley, P.; Kramer, J.
submitted to the EMBL Data Library, May 1997

A/Description: The sequence of C. elegans cosmid R155.

A/Reference number: Z18321

A/Accession: T15281

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-486 <GB>

A/Cross-References: EMBL:AF003390; NID:G2088866; PID:G2088866; PIDN:AA854271.1; GSPDB:GN

A/Experimental source: strain Bristol N2; Clone R155

C/Genetics:

A/Gene: CESP.R155.1

A/Map position: 3

A/Introns: 27/1; 127/2; 171/3; 203/2; 289/2; 324/2; 337/2; 439/3

Query Match	29.1%; Score 749.5; DB 2; Length 486;
Best Local Similarity	32.8%; Pred. No. 5.7e-52;
Matches 157; Conservative 95; Mismatches 203; Indels 23; Gaps 4;	
Qy	31 LATSLSAQSEALRLITIFLGYPPALFYRHYLFYKETYLLHLPHFTTGLSIAYFNGNQL 90
Db	7 LSEVTSASEDALRLISLVLAGYPLAVVHRTFFYKPAQOHOLFVIVGLSLMFNCCSSV 66
Qy	91 YHSLCTIVLQFLRLRMGRITTVLTFFCFQMAVLLAGYVYTAQNDIKMTMPHCVLTL 150
Db	67 IHPILSTFGAFFTTFNAGTDASTYAAHIVFLGHLIGVWFHEDTDYITWTTPFCIMTL 126
Qy	151 KLIGLAVDFPDGQDQNSLSEQOKYAIRGVPSLLEVAGSFYGAFLVGFQFSNMHYMK 210
Db	127 RFLGLVWDVVDGAGKPRHLKRPDQKLTISDKPGLLEIAAGLFGQGLVGPQFLSLKFRS 186
Qy	211 LVQGLIDIDPKLINS-IIPALRSLGLFLYLVGYTLLSPHITEDYLLTDDYDNHPWFR 269
Db	187 FVNDWDVSDGQPPKSAFLPSIGRFLAGCTYVVLHQGQFWIPQYFNSDAVNNLSFFWR 246
Qy	270 CMWMLWGKFLVLYKYVCMVTEGVCLTGLGFGFGEKQAKKADACANMKWMLFENNPR 329
Db	247 WSWVTLWFLRLMXYCAMMLITBGSATLSGLGHNKQABEGNDRDGVRDLHITKWTGHD 306
Qy	330 FTGTIASFNINTAMVAVR-----YIFKRLKFLGNKELSGQLPLFLALHGL 376
Db	307 YNSVSEFNGCTINFPAQNQFAIFNIFATKIHHRRLRWNNKLASGVITLSYALIMHG 366
Qy	377 HSGYLVCQMFLLIVVERQARLIQESPTLSKLAATVLAQPFYVVOQTTHLFWGYSM 436
Db	367 HLGYFLFGVELGCVQANQDLVALIKRTIPQMSSEAIKSPISRPFIWIFGKLTISYMGFAF 426

QY 437 TAFCLFTWDMKMLKYKSIYFLGHIFP-----LSLFLTP-YTHKAMVPRKELKK 485
 DB 427 LMFGLIKTKYVIGPKSYIFIGFIITYFTVPIIHLVLRVPRHRKKAALAEKPEEVK 484

RESULT 2

G96655
 unknown protein, 29405-27288 (imported) - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: G96655
 R:Theologian, A.; Becker, J.R.; Palm, C.D.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Author: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Malt, R.; Marziali, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Author: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MIMD:21016719; PMID:11130712
 A:Accession: G96655
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-465 <STO>
 A:Cross-references: GB:AB005173; NID:g6598844; PIDN:AAPI8699.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F16M19.14
 A:Map position: 1

Query Match 15.8%; Score 407; DB 2; Length 465;
 Best Local Similarity 25.4%; Pred. No. 9.8e-25;
 Matches 123; Conservative 90; Mismatches 227; Indels 44; Gaps 12;

QY 26 LSLINKATSLGASEQALRLIISIFLGYPALFYRHYLHFTFTGLSIAYFN 85
 DB 4 LDMNSMAISIGVAVLRFLLCFVATIPISFLMR---FIPSLRGKHIYAAAGAFSLYS 60
 QY 86 FGNOLYHSLCIVLOFLIRLMGRITTA-----LTFQOMAYILAGYTTATGNY- 137
 DB 61 FG-----FSSNLHFLVMTIGVASMAYRPLSGFTFFPLGFAVILGCHVFMGDMK 113
 QY 138 --DIKWTMPCVTLTKLIGLAVDFPDGKDONSLSSEQKVAIRGVSFLLEVAGSYFPG 195
 DB 114 EGGISTGALMVLTLKVISCSININDGMKEGLRAQKKRLIQMPSLIEYFGCLCG 173
 QY 196 AFLVGPQFSNMHYMKLVQGE---LIDIPKINSIIPALKRLSLGLFVLGYTLSPHIT 252
 DB 174 SHFAGPVFEMKDYLEWTEKGIWAVESEKGRSPGAMIRAVFOALCALVLYLVPDP 233
 QY 253 EBYLLTEVDYDHPFPRCMYMLMGKFLVYKVTCLVTGVCILTGCGNF--EKKK 310
 DB 234 LTRFTEPVYQEWGFLKRFQYQYAGFTARKKYFIWISSEASIIISGLQFSGMTDETQK 293
 QY 311 AKMDACAMKVMFLFTNPRFTGTIASFNINTAMVARYIFKRLKFLGKNE-----LSQG 364
 DB 294 AKMDAKAVDILIGVELAKSAVOIPLFMNIOVSTMLRHVYERIVKRGKAGFOLLATOT 353
 QY 365 LSLFLALMGLHSGYLVCQMEFLIVYERQARLIQ--SPTLSKTAITVLOPFYVLV 423
 DB 354 VS-----AVWHGLYPGIIFVQSAIMIDSKAIYRQQAIPPKMMLRNVLVLLINFLTV 409
 QY 424 QOTIMLPMGYMTAFCLFTWDMKMLKYKSIYFLGHIFLSLFLIPYTHKAMVPRKEL 483
 DB 410 -----VVLNYSVGFMVLSEHTEVAFKSVYVIGTIVIAVL-LLSYL-VVKKVPKPT 461
 QY 484 KQME 487
 DB 462 KKE 465

RESULT 3

hypothetical protein SPBC16A3.10 - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: J39542
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
 submitted to the EMBL Data Library, February 1998
 A:Reference number: Z21862
 A:Accession: J39542
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-509 <WOO>
 A:Cross-references: EMBL:AL021748; PIDN:CAA16861.1; GSPDB:GN00067; SPDB:SPBC16A3.10
 C:Genetics:
 A:Gene: SPDB:SPBC16A3.10
 A:Map position: 2

Query Match 15.7%; Score 406; DB 2; Length 509;
 Best Local Similarity 25.1%; Pred. No. 1.3e-24;
 Matches 121; Conservative 86; Mismatches 206; Indels 70; Gaps 13;

QY 25 ELSINKATSLGASEQALRLIISIFLGYPALFYRHY-----LFYKETYL 69
 DB 6 DIPFVPSFLGVHPDQKLLFCFLSAVFPFAGILKRLSPAWIRNLPSISGLF----YL 61
 QY 70 IHLFTFTGLSIAYFN-----FGNOLYHSLCIVLOFLIRLMGRITTAVALTTFQOMAY 124
 DB 62 IGVHHLVGVVLVLEFDALFTYFAAFYSSSMPWIIFTV--ILGHTFSSHVIRYIY---- 115
 QY 125 LLAGYVYATGNYDICKMTMPCVTLTKLIGLAVDFPDGKDONSLSSEQKVAIRGVSFL 184
 DB 116 -----PSNTDI--TASQVLCMKLTARMSVYDERLSLSSTYQKRALKRTVIT 165
 QY 185 LEVAGFSYFGAFVLGPOFSNMHYMKLVQ---GELIDIPK--IPNSIIPALKRLSLGL 238
 DB 166 LYFLGVVFPFSLVGLGAFVDYDERFTLSMPFLADPYEKQITPHSLPALGCMWGL 225
 QY 239 FYLVGYTLLSHRITEDYLDYDHPFPRCMYMLMGKFLVYKVTCLVTGVCILT 298
 DB 226 LMLILFITGSSIYPLKFLITPKFASSPILLYGVYCIATFAVMKYYGAWELSDACILS 285
 QY 299 GLGPNFPEKKAKMDACAMKVMFLFTNPRFTGTIASFNINTAMVARYIFKRLKFLGN 358
 DB 286 GIGVGLDSSKHPKRDYKATIDPKFPAADIKCALEAMNNTKMLNRYILRYAKKXK 345
 QY 359 KE-LSQGLSLFL-ALMGLHSGYLVCQMEFLIVYERQARLIQ-----ESPTLS 408
 DB 346 RPFKSTLSTFTVSAMHGVAGYLLTFVSAFIQTVAKYTRRHVRPFLKPDMTF--- 402
 QY 409 KLAATTVQPF---YVLVQOTIMLPMGYMTAFCLFTWDMKMLKYKSIYFLGHIFLSL 465
 DB 403 -----GPKRYVDVIGVATNLSLYLIIISFLNLKESIHWMKELVFIIVHILIA 454
 QY 466 LPI 468
 DB 455 LAV 457

RESULT 4

probable membrane protein YOR175c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein OJ635
 C:Species: Saccharomyces cerevisiae
 C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 19-Apr-2002
 C:Accession: S67067; S67063
 R:Hughes, B.; Pohl, T.M.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67067
 A:Accession: S67067
 A:Molecule type: DNA
 A:Residues: 1-619 <HUG>

A:Cross-references: EMBL:Z75083; NID:g1420424; PID:e253056; PID:g1420425; MIPS:YOR175c
 A:Experimental source: strain S288C
 R:Borderline, R.; Camases, A.; Madan, A.; Martin, R.P.; Poch, O.; Tarasov, I.A.; Winsc
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67032
 A:Accession: S67063
 A:Molecule type: DNA
 A:Residues: 270-619 <BOR>
 A:Cross-references: EMBL:Z75083; MIPS:YOR175c
 A:Experimental source: strain S288C
 C:Genetics:
 A:Cross-references: SGD:S0005701
 A:Map position: 15K
 C:Keywords: transmembrane protein
 F:53-69/Domain: transmembrane #status predicted <TM1>
 F:461-477/Domain: transmembrane #status predicted <TM2>

Query Match 15.7%; Score 406; DB 2; Length 619;
 Best Local Similarity 26.2%; Pred. No. 1.6e-24;
 Matches 133; Conservative 79; Mismatches 238; Indels 58; Gaps 17;

OY 28 LNKATSLGASGQRLRLISIFLGPALFVRH-----LFYKETYLI--HLFHTPTGLSI 81
 DB 9 LTKITITNGIDSFTRKVAICLIGSFPLNALRKRIPEKRIKLCCTISMSPYLF----- 63
 OY 82 AYFNGNOLYHSLICIVLQFLIRLMGRITVAVLTTCFQMAVLLAGYVYATGNYDIKW 141
 DB 64 GVLMVSGFRLFLSTMTFTYLISRYRSPKMPHL-NPFWVGHILAINIHQFLNEQOT 122
 OY 142 TM---PHCVTLTKLIGLAVDYFDG---KDQNSLSSEQKVAIRGVPSLLEVAGSYF 193
 DB 123 TVDITTSQMVLAAMKLTSFAMSYDGSCTSESDFKLTGHQKSRABRGHPPLKFLAVAF 182
 OY 194 YGAFVGGQFQFMNHNKLVQBELI-DIP-----GKIPNIIIRALKSLISGL 238
 DB 183 YSTLTGFSFYADDSWLNCEMFRLDPESKKPMRHHGERRQIPKNGKALMKVVGSL 242
 OY 239 FYLVGYTLTSLPHITEDYLTED-YDNHPEFRCMYLIMGFVLKYVTCMLVTEGVCIL 297
 DB 243 AMMLISTIGMGMHFPKYVLYDKGFPTRSFIRIHLPLFLGFIHRKYAAAMTISGSCIL 302
 OY 298 TGLGNGEERK-GKAKMDACANMKWLFETNPRFTGTIASFNINTANAVARIYRRLKFL 356
 DB 303 CGLGNGYDSKTQKIRMDRVNRIDIMVETAQNTREMEAMNMNNKMLKYSVLYRTKK 362
 OY 357 GNKELSGQLSLFL--ALMHGHSGLVLCFQMEFLIYVERQAALIOES--PLTSKLA 412
 DB 363 GKKGFSRSTLFTPLTSAFMHGTRPGYVLTFTAGALY---QTCGKIYRNRPPIFLREDG 418
 OY 413 ITVL--QPFYLVQOTIHWLFMGYSMTAFCLFTMDKMLKYKYSIYFLGHIF--FLSLFL 468
 DB 419 VTPLESKITVLDVGLYAKLAAGVAVQPIIIDLKPSLAWGVSIFYVYHIIIVASFLFLR 478
 OY 469 LPY-----HKAVNPRK--EKLKME 487
 DB 479 GPYAKVTEFESKQPKKEIFIRKQKLE 506

RESULT 5
 T15839
 hypothetical protein C54G7.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 15-Sep-2000
 C:Accession: T15839
 R:Du, Z.
 submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid C54G7.
 A:Reference number: Z18416
 A:Accession: T15839
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-480 <DUG>
 A:Cross-references: EMBL:U04010; NID:g1065453; PID:g1065454; PIDN:AAA81391.1; CESP:C54G7

C:Genetics:
 A:Gene: CESP:C54G7.2
 A:Introns: 27/3; 76/2; 102/2; 185/3; 283/1; 384/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein C08F8.4

Query Match 14.9%; Score 383.5; DB 2; Length 480;
 Best Local Similarity 24.1%; Pred. No. 7.5e-23;
 Matches 112; Conservative 98; Mismatches 221; Indels 33; Gaps 13;

OY 28 LNKATSLGASGQRLRLISIFLGPALFVRH-----LYLHLEHTPTGLSIAYFN 85
 DB 14 LEPLAETIGLEADRVNVLISLACFISYVRKTCSLQVNRQVTVPAVAGIGIVFPC 73
 OY 86 FGNOLYHSLICIVLQFLIRLMGRITVAVLTTCFQMAVLLAGYVYATGNYDIKW 144
 DB 74 FGRAIKHLNALAGSYAIVMFAPTQVHKVILFWSMGLFPIHCYRAILLESYLDVTCF 133
 OY 145 HCVLTCLKLIGLAVDYFDG-KDQNSLSSEQKVAIRGVPSLLEVAGSYFYGAFVGPQF 203
 DB 134 IMVAVERATWMAFNLKDGKADQSKLTEEQKRESLKEIPSLIFEMSFNFQVLTGFAN 193
 OY 204 SMNHYMLVOGE--LIDIPKIPNSIIRALKRLSLGLFYL-VGYTLTSLPHITEDYLTED 260
 DB 194 NYDYIKFLDBKHLVADKHGLPSPGTAAAMKRYQSIFFLAIVTLTGKRYVED-VGTOP 252
 OY 261 YDNHPE--WFRCMYLMIGKFLVLYKYVTCMLVTEGVCILTGLFNGFEKAKMDACAN 318
 DB 253 YFALPFPQMFWMWITTF--FIRCAVYFANVPAICMGSFGSGYDKENAEKLCITN 310
 OY 319 MKWLFETNPRFTGTIASFNINTANAVARIYRRLKFLGNKELSGQLSLFLALMHGHS 378
 DB 311 VLPYQVEMAQSLKETLDGMNIGTFWLRKGYER---APKSIPTVATYTLASAVHGVSI 366
 OY 379 GYLVCFQMEFLIYVERQAALIOES-----PLTSKLAATVLOPFYLVQOTIHWLFM 432
 DB 367 GYVMAFTCGFLTY---AAQTFRRSMWRFLDDHNKFAVDI--FSFIISK---IAL 415
 OY 433 GYSMTAFCLFTMDKMLKYKYSIYFLGHIFPLSFLIYPIYIKAM 476
 DB 416 AYATVSPVTMMYPAVDVLRKYVFIPIHVAIVAFALPKPFKPL 459

RESULT 6
 T19097
 hypothetical protein C08F8.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2000
 C:Accession: T19097
 R:Harris, B.
 submitted to the EMBL Data Library, May 1996
 A:Reference number: Z19073
 A:Accession: T19097
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-540 <WIL>
 A:Cross-references: EMBL:Z73103; PIDN:CAA97423.1; GSPDB:GN00022; CESP:C08F8.4
 A:Experimental source: clone C08F8
 C:Genetics:
 A:Gene: CESP:C08F8.4
 A:Map position: 4
 A:Introns: 35/3; 82/2; 108/2; 146/3; 210/3; 309/1; 344/3; 411/3; 462/2; 515/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein C08F8.4

Query Match 11.5%; Score 297; DB 2; Length 540;
 Best Local Similarity 25.0%; Pred. No. 6.6e-16;
 Matches 114; Conservative 71; Mismatches 171; Indels 100; Gaps 22;

OY 75 TFGTSLAYENFGQOLYHSLICIVLQFLIRLMGRITVAV-----LTFPCQMAVYL 126
 DB 69 TSVGLIFTFYCYGAIAH-----LFINGFSYILMWSVPPQVHKSVFAPAMGYLV 119
 OY 127 A-----GYVYATGNYDIKMTNPHCVTLTKLIGLAVDYFDG-KDQNSLSSEQKVAIRGV 181

[illegible]

```

RESULT 7
T20899
hypothetical protein F14F3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision: 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T20899
R:Kershaw, J.
submitted to the EMBL Data Library, June 1995
A:Reference number: Z19343
A:Accession: T20899
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-480 <WIL>
A:Cross-references: EMBL:Z49937; PIDN:CAA0165.1; GSPDB:GN00028; CESP:F14F3.3
A:Experimental source: clone F14F3
C:Genetics:
A:Gene: CESP:F14F3.3
A:Map position: X
A:Introns: 25/1; 79/1; 102/2; 145/3; 192/1; 241/3; 273/2; 331/3; 385/2; 420/2

Query Match          9.5%; Score 244.5; DB 2; Length 480;
Best Local Similarity 21.3%; Pred. No. 8.7e-12;
Matches 96; Conservative 91; Mismatches 184; Indels 79; Gaps 16;

QY 78 GLSIAYFPGNQLYSL---LCIYLOGLILRLMGRTTVAHLPFCFOMAYILAGYYTA 133
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 70 GFMSLFLIGRKIVYSLGICSAISIQ-LNANKSTPLVPLTPTTYM-FVRFAHYLP 127
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||

QY 134 TGNVDIKMTMPCVLTLLIGLAVDPDG---GKDONSSEOOKYAIRGVPSLLEVAGF 190
   :||::||::||::||::||::||::||::||::||::||::||::||::||
DB 128 VNEVASHTNVLIQILITRLRIGITPBEADAMVHKSDN-----PTRKRYLTPLTEKRAY 182
   :||::||::||::||::||::||::||::||::||::||::||::||::||

QY 191 SYFYCAFLVGQFSNNHMKLVQGLDIPCKIPISITPAK--RLSLGFLYLVGYTLL 247
   ||::||::||::||::||::||::||::||::||::||::||::||::||
DB 183 FYHFGELTGYTYDQ-----MLDISONPLIKSWPPTLEVKSRFVRLIMSVPVVIT 234
   ||::||::||::||::||::||::||::||::||::||::||::||::||

QY 248 SPHTIEDYLTLEDYDNHPEFMRFCMMLIMGKVLAK--YVMCMLTGEGVCLITGLG---- 301
   ||::||::||::||::||::||::||::||::||::||::||::||::||
DB 235 NHYFPLDLIRSDALWEVFFRLRYVAL--IFVVKTYYSMAIAESICVILGITYPA 292
   ||::||::||::||::||::||::||::||::||::||::||::||::||

QY 302 -----FNGFE-----EKGAKMDACANNKWLFTETNPRTGTIASFNINTNAA 344
   ||::||::||::||::||::||::||::||::||::||::||::||::||
DB 293 ASNPRTIMGPTDLNAFDLKTRENIEMSSDAIVNLDIPKEVPSDGRDGMKAMNRSVQTM 352
   ||::||::||::||::||::||::||::||::||::||::||::||::||

QY 345 VARIYFKRLKFLGNKELSEQSLFLALMHGHSGLYVCFQMEFLIVVERKQARLQES 404
   ||::||::||::||::||::||::||::||::||::||::||::||::||

```

```

Db      35 LALVHSRKYKRV-----TMTLSVAWHGTYAGYFMSFGVAMCAIIIEVIFKLY-- 404
QY      405 PTLSTKLAATVLPQEPFYIQQOITIHMLFMGYSM--TAPCLFTWDMKMLKYVSIYELGHI-- 460
Db      405 -----PDVETGVAKPKFRILYTHHTIRCGCFEMLATGFL-----KNADVHIFMS 450
QY      461 ---FTLSLLFILFYIHKAMVPRKELKAKME 487
Db      451 SIYVWLPCLCPFIYSASIKSPKQAQSE 480

```

RESULT 8
T27912
hypothetical protein ZK550.1 - *Caenorhabditis elegans*
C|Species: *Caenorhabditis elegans*
C|Date: 15-Oct-1999 #sequence_rev|version 15-Oct-1999 #next_change 06-Oct-2000
C|Accession: T27912
R|Basham, V.
submitted to the EMBL Data Library, November 1996
A|Reference number: Z20438
A|Accession: T27912
A|Status: preliminary; translated from GB/EMBL/DBJ
A|Molecule type: DNA
A|Residues: 1-366 <RTL>
A|Cross-references: EMBL:Z82287; PIDN:CAM05316.1; GSPDB:IGN00022; CESP:ZK550.1
A|Experimental source: clone ZK550
C|Genetics:
A|Gene: CESP:ZK550.1
A|Map position: 4
A|Intons: 35/3, 82/2, 108/2, 146/3, 191/3, 266/3
C|Superfamily: *Caenorhabditis elegans* hypothetical protein C08F8.4

	Query Match	6.1%; Score 158.5; DB 2;	Length 366;
	Best Local Similarity	21.6%; Pred. No. 4.5e-05;	
	Matches	78; Conservative	57; Mismatches 111; Indels 115; Gaps 16
Oy	78 GLSIAYFNFNGQLYHSLLCIYLQFLILRLMGRITTAVL-----TTFCFOMAYL-LAGGY	130	
Dd	72 GFILFYFCYGNEMIAHFINGFGSYLLM-----ISVLPKHYNHKAVFSFAMAYLFLVHPY	124	
Oy	131 ---YTATCNVDIKMTMPHCIVLTKLIGLAVDYPDGG-DONSLSSEQQKAIKRVPSLIE	186	
Dd	125 RWIYOETINLEFTSGSMWMAVGKITLSSAIT--DGMREDDLALNSGQRDAVNEIPSLID	182	
Oy	187 VAFGSYVFAGTLVYGQFSNMNY-----MKLV-QGEGLIDI PKGI PMSII PALK-RLSL	236	
Dd	183 FASWFAPAQSVIIGP---TNHYSNMSAVLDLKLPKFERGTGRPFSTISTVEKFKAVALIL	239	
Oy	237 GLPYLVGYTLLSPHTTEDYLLTEDDYNHPFWRCMYMLIMWKFVLKYVTCLVTEGVCI	296	
Dd	240 SSFCSALYSLPLPIPTKYPTISE-----YNLL-----	266	
Oy	297 LTGAGFNGFEKGAKKADACANMKVMYLEFTNPRTFGTIASFNINNNANVARIRIFKGLFL	356	
Dd	267 -----STNCNVW-----TDAAWLRRIVERV----	286	
Oy	357 GNKEISQGLSLLFL--ALMHGISGYLVCFQMEFLIVVEROARLIIESPTLSKTAATL	414	
Dd	287 ---EDPRKVIVAVYTGVAHHGLAVEYIUSFLTSAIFTLAKVGSITLFO---AICMAANT	339	
Oy	415 V 415		
Dd	340 V 340		
 RESULT 9 T05909 membrane protein YOR175c homolog - barley (fragment) CSpecies: Hordeum vulgare (barley) CDate: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 08-Oct-1999 CAccession: T05909 R.Hess, W.R.; Golz, R.R.; Boerner, T. Plant Sci. 133, 191-201, 1998			

A:Title: Analysis of randomly selected cDNAs reveals the expression of stress- and defense-related genes in Arabidopsis thaliana
A:Reference number: Z15411
A:Accession: J05909
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-935 <HES>
A:Cross-references: EMBL:AJ000229; NID:e1055245; PIDN:CA03955.1; PID:e1312739
A:Experimental source: cv. Halsea, leaf

Query Match	5.7%;	Score 147;	DB 2;	Length 135;
Best Local Similarity	25.5%;	Pred. No. 0.00012;		
Matches 40;	Conservative 28;	Mismatches 45;	Indels 44;	Gaps 7;

```

QY      292  EGVCLITGFGFGEFEEKG--KAKMPCACAMKWLFEFTNDRFGTIA-----SEFININAMV 345
Dd      1  EAAIIISGLGFGTMSDSSPPKAKMBRAINVDLGE---LAGSAOLPLKMNIOVSTWL 56

QY      346  ARYIFKR-----LKFLGNKEISQGLSILFLAMHGLSHGYVCFOMEFLIV----- 391
Dd      57  RYVYVERLLQKKKKGFLDLSGTQYS-----AIMHGLVPGYIMFVQASALMNGSK 108

QY      392  IVERQARLIOESPILSLAATVLOPEYLVQOITH 428
Dd      109  VIKRWQAVKQRP-----PHIVPTKLILH 133

```

RESULT 10
AE2313
hypothetical protein alr4060 (imported) - *Nostoc* sp. (strain PCC 7120)

A:Note: Nostoc sp. strain FCC 7120 is a synonym of Anabaena sp. strain FCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AE231
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Saezoto, S.; Watanabe, A.; Iriyuch
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840

Query Match	4.5%	Score 116.5;	DB 2;	Length 477;
Best Local Similarity	20.0%;	Pred. No. 0.13;		
Matches	85;	Conservative 62;	Mismatches 177;	Indels 101;
				Gaps 17

```

QY      63 FYKKEYYLLHFFTFPGLE-----SLAYRFGNOLYHSLC-----IVLOFILRLM 107
Dh      9 FFFYRYVAKSLMLMGISIDFVAGLMAISLLFVNASHSSFAIFVCEIIFNVVMWMLRQ 68
QY     108 GRITFAVLTFPCFQMAUYLLAGYUUTAGTGVYDIKWTIRPHCVLLTKLIGLVADVEDGKQD 167
Dh     69 GMAKQAIATTVIIDIILAIYKYL---NFPE-----DVGLVLP----- 106
QY     168 SLSSQKAYAIRGVPSLLEAVGAFSYFYGAFLVGPOFSMMHYKLVQGEIIDIPIKINSI 227
Dh     107 GLADWQSKSIPGMSISIPGLSFYUQMAFVUDSYTSRKRAIGALDYLNVSPFFQYV 166
QY     228 IPAKRLSLGLFVLYGYTLLSHITEDYLLTEBYDNRHPRCMMWLLMGKPYLYKXVY- 286
Dh     167 AGPIERRG-DLF-----DOI-ESFPFKETYPDPEGFRLSLGLEKFLVADANIS 215
QY     287 -----CWVTEGVCILT-----GLSFNGEENGKAKMDACAMMKVLLFETNREFTG 332
Dh     216 YIKLDQAQNPMLVMPFAPFLTQIYVDFGYSIALGLAKF---LGINLTINFLAYYTSQ 272
QY     333 TIASF-----NITNMVAVARIYFRKLKFLG-NKELSGLSLLFLAL---WHGLSHGYLVCF 384

```

Db 273 SINEFRRNRHVTLSIFWERYVF--LELMGKKKKQKAPFLFLFTLTLSGFHMGAAINPVFWG 33.0
 Qy 385 QMEFLVIVIERQAAALIQSPFLSTLKAATLYLQFY-----YLVQOOTHLEFMGYSMT 43.7
 Db 331 AYNGELVIVIRIAGR-----PYYKVGQGRFLMFQFISMALTFSSII 37.2
 Qy 438 AFCLPF 44.2
 Db 373 LGCLPF 37.7

RESULT 11
C69451

C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: C69451
R:Klink, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
R: Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
G: Gloeck, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
N: Nature 390, 364-370, 1997
A:Authors: Uitterlinden, T.; Cottrell, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; PMID:98049343; PMID:9389475

Query Match 4.5%; Score 116.5; DB 2; Length 736;
 Best Local Similarity 21.0%; Pred. No. 0.22; Indels 155; Gaps 27;
 Matches 108; Conservative 69; Mismatches 173;

```

QY 11 VVVLAVLAGVLOSQFQSLNKLATSLGASQALRLIISTFLGVPFA----LFPRH----- 60
Db 17 TWIGIAGMIGAGVPEAL-----TGIAGIAGAPAIITLAFLEPNIITLITGLATAELGSAWP 70
QY 61 -----YLFYKETYVLIHLFHTFTGTLIAAFNFGNQLVHSLTICVLQFLIRLMGRITTAVL 115
Db 71 QAGGYVLIWKEA-----MGDIAGGPMAGWIDMA--AHTIACAL-----YA 107
QY 116 TTFCFQMAVYLLAGYYTATGNYDIKMTMPCVYLTKLIGLAVDYF-----DGKQONSL 169
Db 108 VTFGAFPAEMLVGF-----VGLPIPH-VLTAKISLSLAVSFLAAYVNYGAKESGFL 157
QY 170 SSEQOKAIRGYPSLLEVAGF-----SY-----FYGAFVGPQFSM-----NH 207
Db 158 GS---LVTVLKLILLVPAFGFIIIMLSYPDWQSFPTFPFGAGVLAAMGLTPIAFEG 214
QY 208 YMKLVO--GELIDIPCK-IPINSIIPALKSLSLCLFVYNGVTLTSLPHITEYLLTEDVDNP 265
Db 215 FEIIVQSEBEVDPEKNIIPKAIIVSL-WVAVAIYILIAESLLGA-VRAVY-----P.263
QY 266 FMPRCMYLWIMGKFAFLVYKYVTQWLVTGECIITGLGPNFEEKGAKWADACAMKWYLFE 325
Db 264 SW---MYLGQLAELSLVKVYADSIIMPLCGMILAG-----GLIS 298
QY 326 TNPREFGTIAS-----FNINTNAWAVARYIFKELKELGNKELSQGSILF-----LA 371
Db 299 TISANNATIVYSSSRVIFALSRG-----YHRSLSAINERTKTPHVAIFPSYIIIAASLA 354
QY 372 LMHGHSGLVLCFQMEFLIVYERQARL-----IQBS-----PTISKLAATVLOPFY 421
Db 355 PIEAVASASAIMFILLFLAVNVTLLILRLRRPDIOSAFRLPLVPLVPAIATAALLAVASYF 414
QY 422 LVQO-----TIHWLFMGYSMTARCLFTPMDK 446
Db 415 LITQVHGEGVFLITIANMFIS---SFPYFAYSE 445

```

RESULT 12


```

QY 26 LSLNKLATSLGSEAL---RLIISFLGYPALFYRHVLFYKETYLIHFTFTGLSIA 82
DB 68 LSLNFKIDGFFDLQSLSMFLVITGVGLLIHIFESWYMRKEGQ--SRFAYTNLFIA 125
QY 83 YNPNQNLVHSLCTVLOFLILRLMKRTITAVLTTCFQMAVILAGYTTATGNDIKWT 142
DB 126 SN-----SVLVADNLFMYLWEGVS-----CSYLLIGFYTELKNN--LCA 167
QY 143 MPRCVLT-----LKLGLADVFDGDKDONSLSSEQKVAIRGVPSELVAGSPYFAGF 197
DB 168 FFAFLTRSDVFLMIGMLIY---REPNSENFQRIK-----LSSFLVENVFYLU---- 215
QY 188 LVGPQSMNMYKLVQVGLDIDIPGKIPIISIPALKSLGLPYLVGYTLSPHITEDYLL 257
DB 216 -----DYITLFL--LGVYIGK--SAQLPQOTWLSDA---MGPTVSLIHNATNV 259
QY 258 TEDYDNHPWFRCMWMILGKFLVLYKYVTCMLVTBVCILITGLGNGFBEKAKADACA 317
DB 260 TA-----GVYLIARTHFLPLITPILVYGL----- 285
QY 318 NKKWMLFETNPRFTGTIASFNINTNMAVARYFKRLFKLNGKLSGLFLAL---W 373
DB 286 -----IGTITLVSSISALVKDI--KRI--LAYSTMSQ--IGYMFALGVKAM 328
QY 374 HGLHSGYLVCQMEFLIVIVERQARLIQESPTLSKLAITVLPFYLVQOTIHLFMFG 433
DB 329 SAAITHLIMHAIFFKALLFLSAGSLIKSCNENKIFPMGGLRQKLPFLY-----ISFTVGG 383
QY 434 YSMTAFCLFTWDMKMLKVYSITFLGHIF 462
DB 384 ASLVSPFLIT-----AGFYSKGNILF 404

```

RESULT 15

```

T33437
hypothetical protein ZC190.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33437
R:Nelson, J., Wohlmann, P.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid ZC190.
A:Reference number: Z21344
A:Accession: T33437
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-547 <NEP>
A:Cross-References: EMBL:AF078788; PIDN:AAC26966.1; GSPDB:GN00023; CESP:ZC190.6
A:Experimental source: strain Bristol N2; clone ZC190
A:Genetics:
A:Gene: CESP:ZC190.6
A:Map position: 5
A:Introns: 40/2; 105/1; 165/3; 254/3; 272/2; 338/1; 387/1; 406/3; 433/3; 507/3

```

```

Query Match 4.4%; Score 113.5; DB 2; Length 547;
Best Local Similarity 19.2%; Pred. No. 0.27;
Matches 95; Conservative 113; Mismatches 184; Indels 103; Gaps 26;

```

```

QY 62 LFYKETYLIHFTFTGLSIAVFNQNLVHSLCTVLOFLILRLM--GRTITAVLTTF 118
DB 45 LFSVSIYLLDISH--GLTFEFLFEKQYVHIYFVSVRGITWRFLEHGYSLQS--LSLF 100
QY 119 CFQMAVYLLAGYVYTTATGNDIKWTMPHCVLTCLKLGLADVFDGDKDONSLSSEQKVAI 178
DB 101 LFSIRLLIASLCPITAVRSFKIRIWPISLGSIIISIGFSY-----TVHMQDMESP 152
QY 179 RGVPSLLEVAGFSYFYGAFLVGPQSMNMYKLV-----OGELIDIPGKIPIISIPAL 231
DB 153 TVAMTLVDISSCIF---LIAMQKLSLSYRSTGVGLTNGRFQSLDV--YTWSRHLIPAV 207
QY 232 KRLSLGLF--YVGVTLTLLSPHITD---YLITBEDYN-----HPFW-----FRC 270

```

```

DB 208 LLASLKEPCIIITWVLTISNFTENQAIIWIA--YNNILNIYAVILPFLVQSTVNFOL 265
QY 271 MYMLIWGKPVLYKYVTCMLVTEGVCILTGLGNGFE-----EXGAKAMDACAMKVMLPET 326
DB 266 IIAKQSPFLPLFSISLFLA-----LTH--GFQIFELIGADAENADITSSMRLYVMSL 319
QY 327 NPFRTGTIASFNINTNMAVARYIFRKL--FLGNKLSGSLFLFLMHLHSGYL--VC 383
DB 320 LHEFAFALHSFSLQWPSLDSKHTLKLBRVITAKQITQSVRNQVLSITGLVYISLSWC 379
QY 384 EQ-----MEFLIV-----YERQARLIQESPTLS-----KLAITVLOPFY-----YL 422
DB 380 YAOIHLVDVFIATITNNLVDMASALVILIISQYSIQNYRKTAGIASLEKRFQISDVI 439
QY 423 VQQT-----IHWLFMGYSMTAFCLFTWD--KMLKVYSIYF--LGHIFPLSLPET 468
DB 440 WQALIPVACIAFLIHFTFV---IVVMLSTLNSRYIOMTSHYFENITSVFTICVPEL 496
QY 469 LPYIHKAVPRKEKL 483
DB 497 VVYKHKRI--KTKL 509

```

```

Search completed: November 21, 2003, 14:16:37
Job time : 23 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: November 21, 2003, 14:03:03 ; Search time 17 Seconds

(without alignments)
1347.177 Million cell updates/sec

Title: US-09-938-803-6
Perfect score: 2580
Sequence: 1 MASSAEGDEGTVALAGVLQ.....ILPYIKRAMVPRKELKME 487

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	121	4.7	615	1	NUOL_BUCAP
2	114	4.4	614	1	NUOL_BUCAI
3	109	4.2	588	1	YA71_SCHPO
4	107	4.1	380	1	CYB_GARNE
5	107	4.1	380	1	CYB_PELGA
6	107	4.1	392	1	HYBB_ECOLI
7	106.5	4.1	537	1	YCUB_SCHPO
8	106	4.1	512	1	COX1_LOCM1
9	105	4.1	380	1	CYB_PREHY
10	105	4.1	521	1	COX1_APIIL
11	105	4.1	545	1	NUSM_ALBGO
12	105	4.1	654	1	NUSM_RHIST
13	104.5	4.1	380	1	CYB_BRALA
14	104.5	4.1	395	1	DLTB_BACVU
15	104	4.0	379	1	CYB_TINMA
16	103.5	4.0	1060	1	NKCL_MANSE
17	102.5	4.0	461	1	MYJIN_HELPY
18	102.5	4.0	514	1	COX1_PELSU
19	102.5	4.0	546	1	Y61B_MYCPN
20	102	4.0	379	1	CYB_URSAM
21	102	4.0	498	1	TLCX_RICPR
22	102	4.0	513	1	COX1_MACRO
23	102	4.0	527	1	NUZM_ACACA
24	101.5	3.9	549	1	COX1_LEIRA
25	101.5	3.9	686	1	NUSM_SCHCO
26	101	3.9	380	1	CYB_PULBI
27	100	3.9	380	1	CYB_PAVMU
28	100	3.9	380	1	CYB_PUFTU
29	100	3.9	501	1	ANKH_BRARE
30	100	3.9	919	1	YKOS_YEAST
31	99	3.8	380	1	CYB_GALBE
32	99	3.8	380	1	CYB_LACVU
33	99	3.8	380	1	CYB_LOPGA

34	99	3.8	380	1	CYB_PUFNA	079225 pufinus na
35	99	3.8	447	1	NU4M_ANOGA	P34852 anopheles g
36	98.5	3.8	380	1	CYB_BUTBU	094wq8 buteo buteo
37	98.5	3.8	380	1	Y420_MERTU	057863 methanococ
38	98	3.8	380	1	CYB_ALERU	036319 alectoris r
39	98	3.8	380	1	CYB_BALRE	033989 baleatica r
40	98	3.8	380	1	CYB_COTJA	P35075 coturnix co
41	98	3.8	380	1	CYB_GRUNI	034607 grus nigris
42	98	3.8	380	1	CYB_PACVU	079216 pachyptila
43	98	3.8	381	1	CYB_GALCU	P34868 galeocerdo
44	98	3.8	514	1	COX1_BOVIN	P00396 bos taurus
45	97.5	3.8	460	1	MYJIN_HELPY	Q92kw7 helicobacte

ALIGNMENTS

RESULT 1
ID NUOL_BUCAP STANDARD: PRT, 615 AA.
AC 08K9X7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NADH-quinone oxidoreductase chain L (EC 1.6.99.5) (NADH dehydrogenase I, chain L) (NDH-1, chain L).
DE NUOL OR BUSG157.
GN Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamas I., Klasson L., Canbeck B., Naeslund A.K., Eriksson A.-S.,
RT Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
RA "50 million years of genomic stasis in endosymbiotic bacteria.";
RL Science 296:2376-2379(2002).
CC -I- FUNCTION: NDH-1 shuttles electrons from NADH, via FMN and iron-sulfur (Fe-S) centers, to quinones in the respiratory chain.
CC Couples the redox reaction to proton translocation (for every two electrons transferred, four hydrogen ions are translocated across the cytoplasmic membrane), and thus conserves the redox energy in a proton gradient (By similarity).
CC -I- CATALYTIC ACTIVITY: NADH + quinone = NAD(+) + quinol.
CC -I- SUBUNIT: Composed of 13 different subunits. Subunits nuoh, H, J, K, L, M, N constitute the membrane sector of the complex (By similarity).
CC -I- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -I- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE OF CHLOROPLASTS OR MITOCHONDRIA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE014091; AAM67725.1; -
CC InterPro: IPR003918; NADhub_oxred4.
CC InterPro: IPR003916; NADhub_oxred5.
CC InterPro: IPR001750; Oxidored_q1.
CC InterPro: IPR001516; Oxidored_q1_N.
CC Pfam: PF00361; Oxidored_q1; 1.
CC Pfam: PF00662; Oxidored_q1_N; 1.
CC PRINTS: PR01434; NADHGNASE5.
CC PRINTS: PR01437; NHOXDRPTASE4.
CC Oxidoreductase; NAD, Quinone, Transmembrane; Complete proteome.
KW TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 32 52 POTENTIAL.
FT TRANSMEM 79 99 POTENTIAL.

```

FT TRANSMEM 113 133 POTENTIAL.
FT TRANSMEM 136 156 POTENTIAL.
FT TRANSMEM 168 188 POTENTIAL.
FT TRANSMEM 247 267 POTENTIAL.
FT TRANSMEM 279 299 POTENTIAL.
FT TRANSMEM 327 347 POTENTIAL.
FT TRANSMEM 372 392 POTENTIAL.
FT TRANSMEM 410 430 POTENTIAL.
FT TRANSMEM 457 477 POTENTIAL.
FT TRANSMEM 491 511 POTENTIAL.
FT TRANSMEM 536 556 POTENTIAL.
FT TRANSMEM 594 614 POTENTIAL.
SQ SEQUENCE 615 AA; 70511 MW; 8873DB38A4B7B1 CRC64;

Query Match
Best Local Similarity 4.7%; Score 121; DB 1; Length 615;
Matches 124; Conservative 67; Mismatches 191; Indels 182; Gaps 32;

QY 20 QSGFQELSLINKLATSL-----GASEQALRLIISIFLYPPALFYRHYLFYKETYLIHLPH 74
DB 62 QKLFMSISINELDIDCSLHLDGSLISMILIGI--GLIHIFSTWYMDKSGY--SRFF 117
QY 75 TTGTGSIANFNQNLHSLCTIVLOFLIRLMGRTTAVLTTCFQMAVYLAGYTYAT 134
DB 118 AYTNIPLIASM-----SLVYADNLFEMYL-GWEIVSI-----CYLLIGFYTKT 161
QY 135 GNYDIKMTWPHCVLTUKLIGLAVDYPDGKDNLSSEQKXAIRGVPSSLFVAGSFY 194
DB 162 NN-----TSCALKAFTTRISDVE-----LIISMFLINKYGTENFOEIKLSNPL 207
QY 195 GAFVGPQPSNMHWML-----VOGELIDIP-----GKIP-NSIIPALKRSLGL 238
DB 208 N---VEDCFDLN-VLTLCGLLGWGSQDLPLHTWLSDMVGPPTPSALIHATWTA 263
QY 239 FVLVGYT---LISPHITEDYLI-----TEDYDNHP 266
DB 264 -VLIARTHFLFLTPKIL--YLISLIGITIFISPSFALVODIKRILAVTMSQIGYMF 320
QY 267 -----W-----FCMVYELMGKFL-----YKYVTGW 288
DB 321 LALGVKAMPAIYVHLIYHAIFFKALFLSSGSLSCNNENKFNLSKVSCKEPLVVS-F 379
QY 289 LVTEGVCIITGLGNGFEKGRKAKMDACAMKMYLFETNPRFTGTIASFNINTNAVARY 348
DB 380 LVGGSALVFPPLITSGFYSKGNILFSLVDGYFNLF-----LIGLCSF--LTSITPFM 432
QY 349 IF-----KRLKFL-GKELSGSLFLALMHGSLGVCVOMELIYVER-QARLI 401
DB 433 IIVIFRSSVSFVFSNKRILAHNLPLLILLFSTMGFYFIIRPLFPVFPVMSKLENGKFL 492
QY 402 OE--SPTLSKLA-----ITVLOPFYVL-----VOQTHLMFVMSMTAFCLFTWDKYL 448
DB 493 YETISSFIFLGLFIYHITWIKOPFFRFLKFKIKIHL-----KFLNGW----- 539
QY 449 KYKSIYFLGHIFFLSLFIPLRYI 472
DB 540 -----YF---DFEYKILFIHPYL 554

RESULT 2
NUOL_BUCAI
ID NUOL_BUCAI STANDARD; PRT; 614 AA.
AC P57262;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NADH-quinone oxidoreductase chain L (EC 1.6.99.5) (NADH dehydrogenase
I, chain L) (NDH-1, chain L).
GN NUOL OR BU164.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Buchnera.

```

```

OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RA MEDLINE=20445173; PubMed=1093077;
RX Shigenobu S., Matsumoto H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
BU Buchnera sp. APS."
RL Nature 407:81-86(2000).
CC -I- FUNCTION: NDH-1 shuttles electrons from NADH, via FMN and iron-
sulfur (Fe-S) centers, to quinones in the respiratory chain.
CC Couples the redox reaction to proton translocation (for every two
electrons transferred, four protons are translocated across
the cytoplasmic membrane), and thus conserves the redox energy in
a proton gradient (by similarity).
CC -I- CATALYTIC ACTIVITY: NADH + quinone = NAD(+) + quinol.
CC -I- SUBUNIT: COMPOSED OF 13 DIFFERENT SUBUNITS. SUBUNITS NUOA, H, J,
K, L, M, N CONSTITUTE THE MEMBRANE SECTOR OF THE COMPLEX (BY
SIMILARITY).
CC -I- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -I- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
OF CHLOROPLASTS OR MITOCHONDRIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AP001118; BAB12882.1; -
CC DR InterPro: IPR003916; NADHUB_oxred5.
CC DR InterPro: IPR001750; Oxidored_q1.
CC DR InterPro: IPR001516; Oxidored_q1_N.
CC DR Pfam: PF00361; oxidored_q1_N.
CC DR Pfam: PF00662; oxidored_q1_N; 1.
CC DR PRINTS: PR01434; NADHDHGNASE5.
CC KW Oxidoreductase; NAD; Quinone; Transmembrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 33 53 POTENTIAL.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 136 156 POTENTIAL.
FT TRANSMEM 168 188 POTENTIAL.
FT TRANSMEM 207 227 POTENTIAL.
FT TRANSMEM 247 267 POTENTIAL.
FT TRANSMEM 271 291 POTENTIAL.
FT TRANSMEM 327 347 POTENTIAL.
FT TRANSMEM 372 392 POTENTIAL.
FT TRANSMEM 410 430 POTENTIAL.
FT TRANSMEM 455 475 POTENTIAL.
FT TRANSMEM 492 512 POTENTIAL.
FT TRANSMEM 533 553 POTENTIAL.
FT TRANSMEM 593 613 POTENTIAL.
SQ SEQUENCE 614 AA; 70881 MW; 18CC2D2FC4FE27B0 CRC64;

Query Match
Best Local Similarity 4.4%; Score 114; DB 1; Length 614;
Matches 94; Conservative 66; Mismatches 165; Indels 124; Gaps 20;

QY 26 LSLINKLATSLGSEQAL---RLIISIFLYPPALFYRHYLFYKETYLIHLPHFTGSLGIA 82
DB 68 LSLINEFKIDPFPFLDGLSLMLFVITGVGLIHIFSSMYMKYKEQ--SRFPAYTNLPLA 125
QY 83 YNFGNQLYHSLCTIVLOFLIRLMGRTTAVLTTCFQMAVYLAGYTYATGNDIKMT 142
DB 126 SM-----SVLYADNLFEMYLGMGVSV-----CYLLIGFYTELKNN--LCA 167
QY 143 MHCVLT---LKLIGLAVDYPDGKDNLSSEQKXAIRGVPSSLFVAGSFYFAGF 197
DB 168 FKAFLITVSDVFLMIGMFLY---REFNSFNQEIKF---LSSFLNVNPFYLL----- 215
QY 198 LVGQPSNMHWMLVQGLIDIPGKI PMSIIPALKRSLGLFVAVGYTLLSPHITEDYLL 257

```

Db 216 -----DITLFL--LIGVICK--SAQPLQOTWSDA---WVGPPYSALHHAATMV 259
 QY 258 TEDYDNHFWFRCMYLMWGKFLVYKYVTCMLTEGVCLTLGFGNFEKGRKAKWDACA 317
 Db 260 TA-----GYLLAKTHFLFLTPGLIYLVGL----- 285
 QY 318 NMKVMLEFETNRFPTGTASFNININAMVARYIFRKLFLGKELSGSLFLFLW---W 373
 Db 286 -----IGTLTLIVSSISALVQKDI--KRI--LAVSTMSQ--IGVWFLALGVKAW 328
 QY 374 HGLHSGVLVCFQWFLVIYVERQAARLIQESPTLSKLAITVLOPFYLVQOTIHWLPMG 433
 Db 329 SAATHTLHMAIFPALFLSAGSLIKSCKEKNIFKMGKRLKQDPLFY-----ISFTVGG 383
 QY 434 YSMTAFCLFTWDMKLKVKYSIYFLGHIF 462
 Db 384 ASLVSPFLIT-----AGFYSKNLI 404

RESULT 3

YA71_SCHPO STANDARD; PRT; 588 AA.
 ID YA71_SCHPO

AC Q09758; Q09025; (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 41, Last annotation update)
 DE 28-FEB-2003 (Rel. 40, Last annotation update)
 GN SPAC24H6.01C OR SPAB21F2.01 in chromosome 1.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=218401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Mablet D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Voickart G., Aert R., Robben J., Grymopoulos B.,
 RA Wellens I., Vanstreel E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesl D., Hilbert H.,
 RA Borzani K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Esler P., Zimmermann W., Wedler H., Wambolt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleux V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rocher M., Gallard C., Tallada V.A., Garzon A., Rhode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Usero D., Barrell B.G., Nurse P.;
 RA "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -1 SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1 SIMILARITY: TO YEAST YGL084C.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>)

CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL; Z54142; CA90845.2; --
 DR EMBL; AL590562; CAC36890.1; --
 DR GenBank; SPAC24H6.01C; --
 DR Pflam; PFO3062; MBOAT; 1.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 53 73 POTENTIAL.
 FT TRANSMEM 120 140 POTENTIAL.
 FT TRANSMEM 160 180 POTENTIAL.
 FT TRANSMEM 192 212 POTENTIAL.
 FT TRANSMEM 301 321 POTENTIAL.
 FT TRANSMEM 344 364 POTENTIAL.
 FT TRANSMEM 374 394 POTENTIAL.
 FT TRANSMEM 445 465 POTENTIAL.
 FT TRANSMEM 466 486 POTENTIAL.
 FT TRANSMEM 513 533 POTENTIAL.
 FT DOMAIN 22 40 SER-RICH.
 SQ SEQUENCE 588 AA; 69089 MW; 757AB685494B8B7A CRC64;

Query Match 4.2%; Score 109; DB 1; Length 588;
 Best Local Similarity 16.9%; Pred. No. 0.98; Mismatches 198; Indels 122; Gaps 18;
 Matches 83; Conservative 89;

QY 37 ASQALRLIISIFLGYPPALFYRHYLYFKET---YLHLFHTPTGSLIAYFNGQLYH 92
 Db 114 APENNNPILLLVIVVFLRVLKVLVPTNTNDELAKNNYRLCFSLLFLLVGVGVY 173
 QY 93 SLDCIVQFILMLMGTTTAVLTTFQGMAYLLAGYYT-----ATGN 136
 Db 174 VLTIALINYLISKLKNSINPLTLTLDISVFFKXFPAYCKRSSLHPGLGFLDQYTG 233
 QY 137 YDKMTWPHCVLTKLIGLVNDY-----PDGKDQNSLS--SEQKVYIRG 180
 Db 234 LE-RWVLENTIMRLRSEFMDYWSLKNSKNTLIFDKDEPTTLTRERVDYCLD 292
 QY 181 VP-SLELVAGSFYGAFLVGPQFSNMHMYKLVOGELIDIPGKIPIISIPALKRLSLGLF 239
 Db 293 EDVNLKNFLYIYFAPLYLGLISFNFMNSQ-----KPT--VSTLKVRNL--- 338
 QY 240 YLVGYTLSSHITBEDYLLTDDY-----DNHPFWRCMYLMWGKFLV-Y 282
 Db 339 -LVAIRFLVCVLTMEFLHAYVTALSKDNNMQYSAVESAMISFVLFM-TWLKLI 396
 QY 283 KYVTCMLVTEGV-----CITLGLGFGNFEKGRKAKWDACA NMKVMLEFETNRFPTGT 334
 Db 397 RLFLRLSLIDIDIPPENIVRCMCNNTSAVG-----WRA----- 430
 QY 335 ASFNININAMVARYIFRKLFLGKELSGSLFLFLWGLH-----SGVLVCFQWFL 389
 Db 431 --WHRSTNRRLIYIYVPLGSGNSHSLINLFIITFVALWMDISWELFAKMWLI-----VL 483
 QY 390 IYVERQARL-----IQESPTLSKLAITVLOPFYLVQOTIHWLPMGYSMTAFCLFTW 444
 Db 484 FILPERLCFMSRTGTLKPYRYISFGAALINITYMICNLIGFVAGIDIGKVLVVSF 543
 QY 445 DKWLKVKYSIYF 456
 Db 544 FLTLKGRKFLYF 555

RESULT 4

CYB_GARNE STANDARD; PRT; 380 AA.
 ID CYB_GARNE

AC Q09203;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cytochrome b.
 GN MTYB OR COB OR CYTB.
 OS Gasteria neireis (Grey-backed storm petrel).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Archaeosauria; Aves; Neognathae; Procellariiformes; Procellariidae;
OC Hydrobatidae; Gaviidae.
OX NCBI_TaxID=79649;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate GBS-1;
RX MEDLINE=99003708; PubMed=9787440;
RA Nunn G.B., Stanley S.E.;
RT "Body size effects and rates of cytochrome-b evolution in tube-nosed
  seabirds";
RL Mol. Biol. Evol. 15:1360-1371(1998).
CC -1- FUNCTION: Component of the ubiquinol-cytochrome c reductase
  complex (complex III or cytochrome b-c1 complex), which is a
  respiratory chain that generates an electrochemical potential
  coupled to ATP synthesis (By similarity).
CC -1- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
  b562) is low-potential and absorbs at about 562, and heme 2 (or BH
  or b566) is high-potential and absorbs at about 566 (By
  similarity).
CC -1- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
  cytochrome c1 and the Rieske protein (By similarity).
CC -1- SIMILARITY: Belongs to the cytochrome b family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF076056; AAC68613.1;
DR InterPro: IPR005798; Cytb_b6_C.
DR InterPro: IPR005797; Cytb_b6_N.
DR Pfam: PF00032; cytochrome_b_C; 1.
DR Pfam: PF00033; cytochrome_b_N; 1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE: PS00193; CYTOCHROME_B_CO; 1.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
  Heme.
KW METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 183 183 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 380 AA; 42686 MW; 929F609033BC58F CRC64;

Query Match 4.1%; Score 107; DB 1; Length 380;
Best Local Similarity 20.8%; Pred. No. 0.87;
Matches 75; Conservative 47; Mismatches 106; Indels 132; Gaps 21;

OY 199 VGPOFSNMH-YMKLVQGLIDIPKIPNSI-----IPALKRLSLGLFVLGVYTLSPHIT 252
DB 1 MAPNLRKSHPLKIMNLSLIDLP--TPSNISAMWNFGSLGLCLVTQILTG-LLLATHYT 57
OY 253 EDYLLTEDYDNH-----PFWFRCMY-----LWGKFLVLYKYVTCW 288
DB 58 ADTLTAFSSVHTGCRNVQYGLINLNLHANGASFFPITCIYHIGRLYYGS---YLYKETW 114
OY 289 LVTBVCVILTGL---GFNGFEKKGAKWDACANKWV-----LFTNPRFTGTIASFN 338
DB 115 -NTGILLTLTMTATAFVGY---VLPW---GQMSFGATVITNLSAIPYIGQITIVM- 164
OY 339 INTNANVARYIFKRLKFLGNKELSGSLFLFLAHGHSGLVCFQMEPLIVIERGAA 398
DB 165 ----AMGFSV-----DNPLTIRFFALHFL-----LPPMAGLTLIHL 198
OY 399 RLIESPTLSKLAATVLO--PF--YLVQOTIHWLFNGYSMTAFCLTWD----- 445
DB 199 TFLHSSGNNPLGLVSNCDKIPFHYFSLKDTLGMFLPLITTLALSPMLGDPENFT 258
OY 446 -----KWLKVKYSIYF-----LGHIFELS---LFIPLYIHKA 475
DB 259 PANPLVTPPHIKPEW-----YFLFAVAILRSIPNKLGVLAALASVLIIFLSPHKS 311

```

```

RESULT 5
CYB_PELGA ID CYB_PELGA STANDARD; PRT; 380 AA.
AC 079218;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cytochrome b
GN MTCYB OR COB OR CYTB.
OS Pelecanoides garnoti (peruvian diving petrel).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Archaeosauria; Aves; Neognathae; Procellariiformes; Pelecanoididae;
  Pelecanoides.
OX NCBI_TaxID=79637;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate PDP-1;
RX MEDLINE=99003708; PubMed=9787440;
RA Nunn G.B., Stanley S.E.;
RT "Body size effects and rates of cytochrome-b evolution in tube-nosed
  seabirds";
RL Mol. Biol. Evol. 15:1360-1371(1998).
CC -1- FUNCTION: Component of the ubiquinol-cytochrome c reductase
  complex (complex III or cytochrome b-c1 complex), which is a
  respiratory chain that generates an electrochemical potential
  coupled to ATP synthesis (By similarity).
CC -1- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
  b562) is low-potential and absorbs at about 562, and heme 2 (or BH
  or b566) is high-potential and absorbs at about 566 (By
  similarity).
CC -1- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
  cytochrome c1 and the Rieske protein (By similarity).
CC -1- SIMILARITY: Belongs to the cytochrome b family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF076073; AAC68630.1;
DR InterPro: IPR005798; Cytb_b6_C.
DR InterPro: IPR005797; Cytb_b6_N.
DR Pfam: PF00032; cytochrome_b_C; 1.
DR Pfam: PF00033; cytochrome_b_N; 1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE: PS00193; CYTOCHROME_B_CO; 1.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
  Heme.
KW METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 183 183 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 380 AA; 42653 MW; 7836F01B6AF17840 CRC64;

Query Match 4.1%; Score 107; DB 1; Length 380;
Best Local Similarity 20.5%; Pred. No. 0.87;
Matches 75; Conservative 47; Mismatches 100; Indels 144; Gaps 21;

OY 199 VGPOFSNMH-YMKLVQGLIDIPKIPNSI-----IPALKRLSLGLFVLGVYTL 246
DB 1 MAPNLRKSHPLKIMNLSLIDLP--VPSNISAMWNFGSLTLITCLITQILTG-----L 51
OY 247 LSPHITEDYLLTEDYDNH-----PFWFRCMY-----LWGKFLVLY 282
DB 52 LAMHYTADTTLAFSSVAHTCRNVQYGLINLNLHANGASFFPITCIYHIGRLYYGS---Y 108
OY 283 KYVTCWLVTEGVCLITGL---GFNGFEKKGAKWDACANKWV-----LFTNPRFTG 332

```

DB 109 LYKWT---NTGVILLTLMATAFVGY---VLPW---GQMSFGATVITNLSAIPYIGQ 159
 QY 333 TIASNTNTNMVARYIFRBLKFLGNKELSGSLFLFLAMHGLSHGVLCQMELVI 392
 DB 160 TLVEW---AMGGSV-----DNPTLTFPALHFL-----LPEMTVG 192
 QY 393 VERQARLIQESPTLSKLAITVLO--PF--YVLVQOTIHWLMFGVMTAFLPTMD--- 445
 DB 193 LSMHILTLHSGSNPLGIYSDCKIPHPYFSLKDLIGFTLMLPLITLALFSPNLG 252
 QY 446 -----KWLKVKYSIYF-----LGHIFLS---LLEFIL 469
 DB 253 DPENTPANPLITPPHIXPEW-----YLFPAVALNSIPKLGVALALASVLILFLA 305
 QY 470 PYIHK 475
 DB 306 PFLHKA 311

RESULT 6
 HYBB_ECOLI
 ID_HYBB_ECOLI STANDARD; PRT; 392 AA.
 AC P37180;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable Ni/Fe-hydrogenase 2 B-type cytochrome subunit.
 GN HYBB OR B2995.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / TGI;
 RX MEDLINE=94292472; PubMed=8021226;
 RA Menon N.K., Chateaus C.Y., Dervartanian M., Wendt J.C.,
 RA Shanmugam K.T., Peck H.D. Jr., Przybyla A.E.,
 RT "Cloning, sequencing, and mutational analysis of the hyb operon
 RL encoding Escherichia coli hydrogenase 2.";
 RL J. Bacteriol. 176:4416-4423(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blactner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 CC -1- FUNCTION: PROBABLE B-TYPE CYTOCHROME.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use. By non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC
 DR EMBL; U09177; AA021590.1;
 DR EMBL; U28377; AA69162.1; ALT_SEQ.
 DR EMBL; AE000382; AAC76031.1; -;
 DR PIR; A65086; A65086.
 DR Ecogen; EG11800; hydB.
 DR InterPro; IPR005614; NtFD.
 DR Pfam; PF03916; NtFD; 1.
 DR Transmembrane; Electron transport; Heme; Complete proteome.
 FT TRANSMEM 12 32 POTENTIAL.
 FT TRANSMEM 35 55 POTENTIAL.

FT TRANSMEM 59 79 POTENTIAL.
 FT TRANSMEM 91 111 POTENTIAL.
 FT TRANSMEM 134 154 POTENTIAL.
 FT TRANSMEM 169 189 POTENTIAL.
 FT TRANSMEM 208 228 POTENTIAL.
 FT TRANSMEM 250 270 POTENTIAL.
 FT TRANSMEM 282 302 POTENTIAL.
 FT TRANSMEM 334 354 POTENTIAL.
 FT TRANSMEM 356 376 POTENTIAL.
 FT CONFLICT 306 307 NV -> KL (N REF. 2).
 SQ SEQUENCE 392 AA; 43574 MW; CECF83746BCABFO CRC64;
 Query Match 4.1%; Score 107; DB 1; Length 392;
 Best Local Similarity 21.0%; Pred. No. 0.89;
 Matches 89; Conservative 50; Mismatches 135; Indels 150; Gaps 21;
 QY 79 LSIAYENFGNQLYHSLCTIYQIPLRLMGFTITAVLTTCFQMAVLIAGYYTA-TGNY 137
 DB 72 LAMAVYVENRGQYHPL---VRPALASLFG-----YSIGSLSTIDVGRY 113
 QY 138 DIKTMHPCVLTLLGLAVDYFDGKDQNSLSEQQ-----KYAIRGPSLLEAVG 189
 DB 114 ---WNLPEF-----YIPGHFNVNSVLEPFAVCMTIYIGWMLERAPALPERLG 158
 QY 190 FSYFGAFLVGPQPSMNMHMLVOGELIDIPGKIIPNSIIPALKRLSLGLFYLVGYTLSP 249
 DB 159 WKV-----SLGRUNKMFFIHALGALLP 181
 QY 250 HITF---DYLLTFDYDHPFW-----FRCMYMLWGFVLKYVTCMLVTEGCIL 297
 DB 182 TMHOSMSGLMSISAGYKVHPLMQSYEMLPFLSLTLAPFMGPSI-----VIEGSLVQ 233
 QY 298 TGLGFNGEEBEGKAKMDACAMKMWLPETNRFCTTISFNINNAMVARYIFRKLKLG 357
 DB 234 AGLRNGNPDERS-----LP---VKLTNTIS-----VLALFIVLRF-G 267
 QY 358 NKELSQGLSLFLALMHGLSHGVLCFQMEFLIVVERQARL--IQSPFLSKLAITV 415
 DB 268 ELIYRDKLSLAFAG-----DFYSVMFIEVLMLFLVLRVANVRDSMLPLSALSA 321
 QY 416 LQPFYLVQOTIHWLMFGVMTAF-----CLF-TWDRMLKV--YKSIYFLGHIFFLSL 466
 DB 322 L-----LGCATWR-LTYSIVAFNPGGYAVFPTMBELLISIGVAIEICAYIVLIRLL 373
 QY 467 FILP 470
 DB 374 PILP 377

RESULT 7
 YCUB_SCHPO
 ID_YCUB_SCHPO STANDARD; PRT; 537 AA.
 AC O59831;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-SEP-2003 (Rel. 41, Last annotation update)
 DE Putative amino-acid permease C965.11C.
 GN SPC965.11C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagske K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbittsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Symmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purrelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas R., Rochet M., Galliard C., Tallada V.A., Garçon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forstberg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Usero D., Barrell B.G., Nurse P.,
 RT "The genome sequence of Schizosaccharomyces pombe.",
 RL Nature 415:871-880(2002).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
 CC -1- SIMILARITY: Belongs to the amino acid permease family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AL023590; CAA19071.1; -
 CC PIR: T41664; T41664
 CC GenBank: SP00065.11c; -
 CC InterPro: IPR002293; AA/re1_permease1.
 CC InterPro: IPR004840; AAC_permease.
 CC InterPro: IPR004841; Permease.
 CC Pfam: PF00324; aa_permeases: 1.
 CC PROSITE: PS00218; AMINO_ACID_PERMEASE_1; FALSE_NEG.
 DR Hypothetical protein; Transport; Amino-acid transport; Transmembrane.
 FT TRANSMEM 52 72 POTENTIAL.
 FT TRANSMEM 76 96 POTENTIAL.
 FT TRANSMEM 157 177 POTENTIAL.
 FT TRANSMEM 183 203 POTENTIAL.
 FT TRANSMEM 219 239 POTENTIAL.
 FT TRANSMEM 267 287 POTENTIAL.
 FT TRANSMEM 318 338 POTENTIAL.
 FT TRANSMEM 362 382 POTENTIAL.
 FT TRANSMEM 393 413 POTENTIAL.
 FT TRANSMEM 462 482 POTENTIAL.
 FT TRANSMEM 537 AA; 59431 MW; C6022537E9AE1CF CRC64;
 SO SEQUENCE

Query Match 4.1%; Score 106.5; DB 1; Length 537;
 Best Local Similarity 20.1%; Pred. No. 1.4; Indels 91; Gaps 17;
 Matches 67; Conservative 58; Mismatches 118;

QY 5 AAGDECTVAL--AGVLQSGFQELSLNKATLSGASEQALRLISFLGPPALFRNHL 62
 DB 222 AAGFGIVSAFYVAGVFSGVSVMT--AAESKNPKKAIPLAV----- 263
 QY 63 FKETYLHLHFTFTGLSIAY-----FNFGQLVSLCYLQFLIRLMGRIT 112
 DB 264 --RQTFW-RILYVFPISISYGTITVAMDPNLSGSKTLSPMTIIMAGNMHAGDPN 320
 QY 113 AVLTFPCFMAYLLAGYVYATGNVD-----IKWTMPCVLLKLG 154
 DB 321 AVILITC--LSSINSIGYISRLVYLMADQMARIKFRVDKRGVPMVAHSVHLEGFDS 378
 QY 155 LAVDYFDGQDNSSLSSEQKAIKRVPSLLEVAGSYFGALVLPQSSNHHM--K 210
 DB 379 I-MNVTG-----AVKAYGYIINLAGS-----AIVTATIIFFVFRFRRW 419
 QY 211 LVQGS-ELIDIPKIPNSIIPALKRLSLGLFYLV--GYTLLSPHTEIDYLLTEDYDNHPW 267

DB 420 VKQYALSDLPFKSLPVEPQLIGVIGIILTLVQGWTFVFPFAGDPV--DAYILBLF 477
 QY 268 FROMMLWKGKFLVLYKY--TCMLVTEVCILTG 299
 DB 478 F-----VTM-----LSYKFKTKTKWVSIEDMDING 503

RESULT 8
 COX1_LOCM1 STANDARD; PRT; 512 AA.
 ID COX1_LOCM1
 AC Q36421;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
 GN COL.
 OS Locusta migratoria (Migratory locust).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Orthopteroidea; Orthoptera; Caelifera; Acridoidea;
 OC Acridoidea; Acrididae; Oedipodinae; Locusta.
 OK NCBI_TaxID=7004;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96139026; PubMed=8587138;
 RA Flock P.K., Rowell C.H.P., Gellisen G.;
 RT "The sequence, organization, and evolution of the Locusta migratoria
 RT mitochondrial genome.";
 RL J. Mol. Evol. 41:928-941(1995).
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1 -
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocytochrome c + O(2) = 4 ferricytochrome
 CC c + 2 H(2)O.
 CC -1- PATHWAY: Respiratory chain; terminal step.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X80245; CAA56527.1; -
 CC PIR: T11467; T11467.
 CC HSSP: P00396; 20CC.
 CC InterPro: IPR000883; COX1.
 CC Pfam: PF00115; COX1; 1.
 CC PRINTS: PR01165; CYCOXIDASE1.
 DR PROSITE: PS00077; COX1; 1.
 DR Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 KW Respiratory chain; Inner membrane.
 KM METAL 60 60
 FT METAL 239 239 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT METAL 243 243 COPPER B (PROBABLE).
 FT METAL 289 289 COPPER B (PROBABLE).
 FT METAL 290 290 COPPER B (PROBABLE).
 FT METAL 375 375 IRON (HEME A3 AXIAL LIGAND) (PROBABLE).
 FT METAL 377 377 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT METAL 377 377 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT CROSSLINK 239 243 1'-histidyl-3'-tyrosine (By similarity).
 SO SEQUENCE 512 AA; 56814 MW; 98031A229B00511C CRC64;

Query Match 4.1%; Score 106; DB 1; Length 512;
 Best Local Similarity 21.7%; Pred. No. 1.4;

Matches 102; Conservative 55; Mismatches 155; Indels 158; Gaps 25;

QY 88 NOLYSLICIVLQFLRLMGRITAVLTTCFCOMAYLAGYATATGVYDK-----140
 DB 50 DQLVVIL-----TAAAFVIFPMVPMIGF-----GNLVPLMIGAPD 90
 QY 141 -----WTMPCVLTLEKLGAVDYFDGKDNLSSEOOKVARIQVPSL-----184
 DB 91 MAFPRMNMNFWLBPSS-LTLLMSSVVD-----NAGAGMTVY-----PLASVIA 136
 QY 185 -----LEVAGFSYFGA--FLVGPQFSNMHYKLVQGLIDIPGKIPNSIIPA 230
 DB 137 HSGASVDIAISLHLAGVSSILGAINFTTAINMNSNMNTLDQTLF-----VMSVATA 191
 QY 221 LKRLSLGLFYLVG-YTLISPHITEDYLITEDYD-----NHPFWF--RCMYMLIM 276
 DB 192 L-LLLSLPVLGALATML---LTDRLNLTSPFDPBAGGDPILYQILFMFGHPEYIIL 247
 QY 277 GKEVLVYKVTW-----LVTEG-VCIITGLFNGFEEKKAKWDACANMKVWL-----323
 DB 248 PGFGIISHIVCOESKIESFGTIGMIVAMLSIGLMGF-----IVMAHMT 293
 QY 324 -----PENTPRPTGTIASFRININ---AMVAVYIKRLKFLGNKELSGLSLFLALMHG 375
 DB 294 VGMDDVTAAVFTSATMIIAIVPTGIKVFSSMATLYGTKEKF--NPPLMALGFIFLFTMG 351
 QY 376 LHSGLVCFQMEFLIVIVERQARLIQSPSLKLAATVLOPFYVLOQTTHLMFGYS 435
 DB 352 LTG--LVLANSLDVLVHD-----TYVVAHFVYLSMGAVFAIMGCI 392
 QY 436 MTAFLCLFTW---DKMLKVKYSIVFLGHIFLSLFI-LPYTHKAMVPRK 480
 DB 393 IOWYLFGLTWNMSKMLKIQFTIMFIG---VNLFFPQHFLGLAGMWR 438

RESULT 9
 CYB_PTEHY STANDARD; PRT; 380 AA.

AC 0793224;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DE Cytochrome b.
 GN MTCYB OR COB OR CYTB.
 OS Pterodroma hypoleuca (Bonin petrel).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Procellariiformes; Procellariidae;
 OC Procellariinae; Pterodroma.
 NC NCB1_TaxID=79640;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate Bonin-1;
 RX MEDLINE=99003708; PubMed=9787440;
 RA Num G.B., Stanley S.E.;
 RT "Body size effects and rates of cytochrome-b evolution in tube-nosed
 seabirds";
 RL Mol. Biol. Evol. 15:1360-1371(1998).
 CC -1- FUNCTION: Component of the ubiquinol-cytochrome c reductase
 complex (complex III or cytochrome b-c1 complex), which is a
 respiratory chain that generates an electrochemical potential
 coupled to ATP synthesis (By similarity).
 CC -1- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
 b562) is low-potential and absorbs at about 562, and heme 2 (or BH
 or b566) is high-potential and absorbs at about 566 (By
 similarity).
 CC -1- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
 cytochrome c1 and the Rieske protein (By similarity).
 CC -1- SIMILARITY: Belongs to the cytochrome b family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)
 CC or send an email to license@ebi.ac.uk).

DR EMBL: AF076079; AAC68636.1;
 DR InterPro: IPR005798; Cytochrome b, c1.
 DR InterPro: IPR005797; Cytochrome b, c1.
 DR Pfam: PF000033; Cytochrome b, c1.
 DR Pfam: PF000033; Cytochrome b, c1.
 DR PROSITE: PS00192; CYTOCHROME B_HEME; 1.
 DR PROSITE: PS00193; CYTOCHROME B_HEME; 1.
 DR Electrophoretic transport; Mitochondrion; Respiratory chain; Transmembrane;
 Heme.
 KM METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 183 183 IRON 1 (HEME B566 AXIAL LIGAND).
 FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).
 SQ SEQUENCE 380 AA; 42668 MW; A8217408D11FB49B CRC64;

Query Match 4.1%; Score 105; DB 1; Length 380;
 Best Local Similarity 21.0%; Pred. No. 1.2; Indels 126; Gaps 20;
 Matches 75; Conservative 46; Mismatches 110;

QY 199 VGPQFSNMH-YMKLVQGLIDIPGKIPNSI-----IPALKRLSLGLFYLVGTYTLSPHIT 252
 DB 1 MAPNLRKSHPLRLKVNNSLIDLP--PSNIGAMNFGSLGICMTQTLTG-LLLAMHYT 57
 QY 253 EDYLTEDYDNH-----PMPFRCKMYLWKGKVL--KYVTCWLVT 291
 DB 58 ADTLTAFSSVHATGRNVQYGLIRNLHANGASFFPICYLHIGRFGYGSYLAKETW--N 115
 QY 292 EGVCLITGL---GNGFEEKKAKWDACANMKW-----LFENPRFTGTIASFRININ 341
 DB 116 TGVMLLTMTATAVGY---VLPW---GQMSFWGATVITNLFAPITGQTLVFW-----164
 QY 342 NAMVAVYIKRLKFLGNKELSGLSLFLALMHGHSGLVCFQMEFLIVIVERQARLI 401
 DB 165 -AMGFSV-----DNPTLIRFPALHTL-----LPFIAGLITLHILFL 201
 QY 402 QESPTLSKLAATVLO--P--YYLVQQTTHLMFGYSMTAFCLFTWD-----445
 DB 202 HESGNNPLGIIVSNCDKIPFHPYFTLKIDLFEMFLPLTTLALFSPPLDGPENFTPAN 261
 QY 446 -----KMLKVKYSIVF-----LGHIFLS---LFIPIYTHKA 475
 DB 262 PLITPPHIKPEW-----YPLFAVAILRSIPNKLGGVLAASVLLIFLVPFLHKA 311

RESULT 10
 COX1_APILI STANDARD; PRT; 521 AA.

AC P20374;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
 GN COI.
 OS Apis mellifera ligustica (Common honeybee).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preyigota;
 OC Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
 OC Apidae; Apis.
 NC NCB1_TaxID=7469;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thorax;
 RX MEDLINE=90136028; PubMed=2559293;
 RA Crozier R.H., Crozier Y.C., Mackinlay A.G.;
 RT "The CO-I and CO-II region of honeybee mitochondrial DNA: evidence
 RT for variation in insect mitochondrial evolutionary rates";
 RL Mol. Biol. Evol. 6:399-411(1989).
 RN [2]

```

RP SEQUENCE FROM N.A.
RC TISSUE=Thorax;
RX MEDLINE=93114603; PubMed=8417993;
RA Crozier R.H., Crozier Y.C.;
RT "The mitochondrial genome of the honeybee Apis mellifera: complete
RT sequence and genome organization.";
RL Genes 133:97-117(1993).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BINETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B.
CC -1- CATALYTIC ACTIVITY: 4 ferrocytochrome c + O(2) = 4 ferricytochrome
CC c + 2 H(2)O.
CC -1- PATHWAY: Respiratory chain; terminal step.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M23409; AAA18476.1; -
DR EMBL: L06178; AAB96799.1; -
DR PIR: A32431; A32431.
DR HSSP: P18401; 1FTF.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1; 1
DR PRINTS: PR01165; CYCOXIDASE1.
DR PROSITE: PS00077; COX1; 1.
KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW Respiratory chain; Inner membrane.
FT METAL 59 59 IRON (HEME A AXIAL LIGAND) (PROBABLE).
FT METAL 238 238 COPPER B (PROBABLE).
FT METAL 242 242 COPPER B (PROBABLE).
FT METAL 288 288 COPPER B (PROBABLE).
FT METAL 289 289 COPPER B (PROBABLE).
FT METAL 374 374 IRON (HEME A3 AXIAL LIGAND) (PROBABLE).
FT METAL 376 376 IRON (HEME A AXIAL LIGAND) (PROBABLE).
FT CROSSLINK 238 242 1'-histidyl-3'-tyrosine (by similarity).
SQ SEQUENCE 521 AA; 59293 MW; 2149417AC981CE64 CRC64;

Query Match
Best Local Similarity 4.1%; Score 105; DB 1; Length 521;
Matches 93; Conservative 61; Mismatches 148; Indels 178; Gaps 24;

QY 88 NOLYHSLCIVLQFLILRLMGRITVAVLTFFCQMAVYLAGYVYRTGNYDIK----- 140
DB 49 DOIYNTIV-----TSHAFLMIFFWMPFLIGF-----GWLPLPLMGSPD 89
QY 141 -----WTMPHCVLTLKLGAVDYFDGKQNSLSSEQKVAIRGVPSL----- 184
DB 90 MAFPRMNNISFWLIPSLMLLSLN-----FYSPGCGWTVYPPLSAYLVHSSDSVDFAI 146
QY 185 -LEVAGFSYFYGA-----FLVGGPQSMNHMYMLVOGELLIDIPGLKPSIIPALKLSLG 237
DB 147 SLHMSGISIMGSLNLMVTIMMKNSMY-----DQSLSPWGVFTAILILNS 196
QY 238 LTVLVG-YTLSPHTEEDYLTEDVD-----NHPWF--RCMTMLWGKRVLYK 283
DB 197 LPLVLAGAIVML-----LFDNRNFTSFPDPWGSDPILYOHLFWFGHBEVYVILILPGGLIS 253
QY 284 VYTCMLVTEG-----VCILTGFGNGFEEKSKAKMDACAMKMWL----- 323
DB 254 HI--VWNEGSKKEIFGNLSMTIYAMLGIFLG-----IWAHMTFTVGLD 296

```

```

QY 324 FETNREFTGTIASFNINTN-----AWVARYIFKRLKFLGNKLSQGLSLFLALWHGLHSG 379
DB 297 VDTRAYFTSATMIIVPFGIKVFSWLATYHSGSKL-----LNISILW--SLG 341
QY 380 YIVCGMEFLVIVYRQARLRQESPTLSKLAIVLOPFY-----LV 423
DB 342 FIMLFTIGTLGIM-----LSNSIDILHDYTVVGHFHYVLSMGAVALI 388
QY 424 QOTIHW--LEMGYSTATCFLFTWDKLVKYSIYELGHIFLSLFLIPYIHKAM-VPRK 480
DB 389 SSFIHWYELITGLNLNT-----KMLKIQIFMTEIG-----VNLTFPGHFLGLMSMR 437

RESULT 11
NUMS_ALBICO STANDARD; PRT; 545 AA.
ID NUMS_ALBICO
AC P48918;
DT 01-FEB-1996 (rel. 33, Created)
DT 01-FEB-1996 (rel. 33, Last sequence update)
DT 01-FEB-1996 (rel. 33, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3).
GN ND5.
OS Albinaria coerulea (land snail).
OC Mitochondrion.
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC Sigurethra; Clausilidae; Clausiliidae; Altopinae; Albinaria.
OX NCBI_TaxID=42349;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96120351; PubMed=7498775;
RA Hatzoglou E., Rodakis G.C., Iecanidou R.;
RT "Complete sequence and gene organization of the mitochondrial genome
RT of the land snail Albinaria coerulea.";
RL Genes 140:1353-1366(1995).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X83390; CAA58296.1; -
DR PIR: SS9143; SS9143.
DR InterPro: IPR003916; NADHUB_oxred.
DR InterPro: IPR001750; Oxidored_q1.
DR InterPro: IPR001516; Oxidored_q1_N.
DR Pfam: PF00361; oxidored_q1; 1.
DR Pfam: PF00662; oxidored_q1_N; 1.
DR PRINTS: PR01434; NADHHDGNASE5.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 545 AA; 61335 MW; B8801FDECTA18563 CRC64;

Query Match
Best Local Similarity 4.1%; Score 105; DB 1; Length 545;
Matches 81; Conservative 42; Mismatches 126; Indels 86; Gaps 19;

QY 4 SAEDEGVVALAGVLSGFQELSLNKLATISGASEQALRLITSFLGYF-AL--FYRH 60
DB 231 SLMGVGSITCLIGSGVALF--EYDKKV-IALSTLSQGVVMYSLSLMLPYALHLTYGH 288
QY 61 YLFYKETVLIHFHTFGLSIAYFNGVQ-----LYHSLLCT-VIQFLRLMGRTI 111
DB 289 AMFKAMLP-----GAGL-IIMNSYGTODLRLLGSLIYSSPTISLISLWCLMGPEP 341
QY 112 TA-----VLTFFCQMAVYLAGY-----YTAGNYDIK----- 140
DB 342 VSSFSYKILILEKMLDNMNCNFTSMFYLGLITLTGMYSIRLMKFLCKWGNNNKPSYCNMS 401
QY 141 WTMPHCVLTLK-----LIGLADVFDGKQNSLSSEQKVAIRGVPSLLEVAGFSYFYGA 196

```

Db 402 WOSKSMFPLAALAVSGLMSYLDSSYMTFSWSTNQNLIMGVLFISIFGIWFKGN 461
 QY 197 F-----LVGP-QFSNMHYMK--LVQGELEIDIPKIPNISIPALKSL-----LGPF- 239
 Db 462 FYPTLMSMMFLGPFPSYNLHTKSLIIMKRIIDISIEPNVMNLMYSSSRWMSLFN 521
 QY 240 YLVGTTLSPHITEDYLTEDYDNHPFWRFCWYMLIW 276
 Db 522 WLTNYMLVLT-----WFLLV-----WLMIMSLIMW 545

RESULT 12

NOSM_RHIST
 ID NOSM_RHIST STANDARD; PRT; 654 AA.

AC P50367;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 5 (BC 1.6.5.3).
 GN NDS OR NAD5.
 OS Rhizopus stolonifer (Rhizopus nigricans).
 OG Mitochondrion.
 OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
 OC Rhizopus.
 OX NCBI_TaxID=4846;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DAOM 148428;
 RA Paquin B., Roewer I., Wang Z., Lang B.F.
 RT "A robust fungal phylogeny using the mitochondrially encoded nads
 protein sequence."
 RL Can. J. Bot. 73:S180-S185(1995).
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; U17011; AAA9062.1; -.
 DR PIR; T14202; T14202.
 DR InterPro; IPR003916; NADHox oxidred.
 DR InterPro; IPR001750; Oxidored_g1.
 DR InterPro; IPR001516; Oxidored_g1_N.
 DR Pfam; PF00361; Oxidored_g1; 1.
 DR Pfam; PF00662; Oxidored_g1_N; 1.
 DR PRINTS; PR01434; NADHDGNASE5.
 DR Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
 SQ SEQUENCE 654 AA; 72818 MW; 22D2E05B8D6E12D CRC64;

Query Match 4.1%; Score 105; DB 1; Length 654;
 Best Local Similarity 21.5%; Pred.No. 2.1;
 Matches 87; Conservative 59; Mismatches 164; Indels 94; Gaps 20;

QY 14 ALAGVLGGFQELSLINKATSGASEQALRLIISIFLGIPPLF-----YRHVLFYKE 66
 Db 289 ATGTGLQNDLKRVIAYSTCSQSG-----LLFVCGLSQYNVALFHLVNHAMFKALLFSA 343
 QY 67 TYLIHLFH-----TFTGLS-IAYNFGNQLYHSLCYLQFLIRLMGRTTAVLT--- 116
 Db 344 GSVIANMDEODLRKFGGLSRLLPFTYSMMVIGSLMPLPL-IGFYSKDIITELAYGH 402
 QY 117 -TFCFOMAYLLAGYYTATGNDYDKMTWCHVLTLLIGLAVDYDGGK-DONSLSSEQQ 174
 Db 403 YFSGNLVYMLASVAAVFTAMYSIR-----SLVLTFLG-----YNGPKINYNNTI--HE 449
 QY 175 KVAIGVPSLLEVAGSFYFG-----AFVGPQFSNMHYMKLVQGELEI 217
 Db 450 ADLMAIAP-LVLAVALSIFFGVYTKDLFVGMGTDFFYNALLFTHPHSI-----LV 498

QY 218 DIPKIPNSI--IPALKSLGLFVLYGVYLLSPHITEDYLLTE-----DYDNHPFWR 269
 Db 499 DTEFGLPWSMKFELPLIGSL-LGTFGVALIYWIPOELPKFKISTYGRGIYFFPNQKYFD 557
 QY 270 CMYMLINKFPLVKKYVTQWLTVEGVCLITGTFGPFGEKGAQKADCAKMYWLPETNPR 329
 Db 558 NIYNMLNKKFLNFGYTTKILDRGAIELVG-----PYGLVNVFKSASNNVSGLDSC-- 608
 QY 330 FTGTIASPINMTNMAVARIYFGRLLKFLGNKELSGQLSLPLALW 373
 Db 609 FIPTYAMTIFNGLLIFTLIF-----FIDPR-----LFWLLIM 642

RESULT 13

CYB_BRALA
 ID CYB_BRALA STANDARD; PRT; 380 AA.

AC P92472; O79415;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cytochrome b.
 GN MTCYB OR COB OR CYTB.
 OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 OX NCBI_TaxID=7740;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96292550; PubMed=9628930;
 RA Spruyt N., Delarbre C., Gachein G., Laudet V.
 RT "Complete sequence of the amphioxus (Branchiostoma lanceolatum)
 RT mitochondrial genome: relations to vertebrates."
 RL Nucleic Acids Res. 26:3279-3285(1998).
 RN [2]
 RP SEQUENCE OF 186-380 FROM N.A.
 RA Gachein G.;
 RT Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: Component of the ubiquinol-cytochrome c reductase
 CC complex (complex III or cytochrome b-c1 complex), which is a
 CC respiratory chain that generates an electrochemical potential
 CC coupled to ATP synthesis (By similarity).
 CC -1- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
 CC b562) is low-potential and absorbs at about 562, and heme 2 (or BH
 CC or b566) is high-potential and absorbs at about 566 (By
 CC similarity).
 CC -1- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
 CC cytochrome c1 and the Rieske protein (By similarity).
 CC -1- SIMILARITY: Belongs to the cytochrome b family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; Y16474; CAA76246.1; -.
 DR EMBL; Y09849; CAA70979.1; -.
 DR PIR; A71390; A71390.
 DR InterPro; IPR005798; Cytb_b6_C.
 DR InterPro; IPR005797; Cytb_b6_N.
 DR Pfam; PF00032; Cytochrome_b_c1.
 DR Pfam; PF00033; Cytochrome_b_N; 1.
 DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
 DR PROSITE; PS00193; CYTOCHROME_B_QQ; 1.
 KW Election transport; Mitochondrion; Respiratory chain; Transmembrane;
 KW Heme.

FT METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
 FT METAL 183 183 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).

```

FT CONFLICT 247 S -> C (IN REF. 2).
SQ SEQUENCE 380 AA; 42639 MW; 514EAB840537E814 CRC64;
Query Match
Best Local Similarity 21.3%; Score 104.5; DB 1; Length 380;
Matches 83; Conservative 48; Mismatches 120; Indels 139; Gaps 23;
200 GPOFNMNHYMKLVQGEIDIPGKIPNSI-----IPALKRLSLGLFYLVTGL-- 246
3 GFLRKHPPLKLVNHSVIDLP--VPSNISVMNFGSLDLCLVSOILTGLFLAMHTAV 60
247 -----LSPITED---YLTEDYDN-HPPFRQMTM-----LWGEFVLYKYVTCWLT 291
61 NLAFFSVAHICRDVNVGWLRLMHNAGASFMEICLYMHIGRGLYGS---YFVETWNI 117
292 EGVCLIT-GLGFNGFEKGRKAKMDACANKW-----LFEINRPFCTGTI-----ASN 338
118 VMLVLTVMATPLG---VLEPW---GQMSFGATVITNLFSAIPFGPDLVQWMLGFS 170
339 INTNMAVARYIFKRLKFLGNKELSGSLFLALMH-----GL-----HSGYL 381
171 VD-NATLRF-FAFHFL--PMIAGLSVHLFLHQTGANNPTGLAGVDKVPFLAYFS 226
382 VCFQNEFLIVVERQARLIQSPYL----- 407
227 YKDVVGFVLLAGLVIALF--SPNLLTDPENYIPANPLVTPVHIQPEWYFLFAVALRS 284
408 --SKLAIVL-----QPPYIVQGTI-----WLFMGYSMTAFCLTMMK 446
285 INNKGGVALMNSIVLFMPFVSHSSRQTSNHPRLAQLFWLM---VVNVLTLTWLG 340
447 WLKVKYSIYFLGH---IFFLSLFLPIY 472
341 GQPEVEPYIFLQASVIFVNIILFPIY 370

```

```

RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maul C., Medigue C.,
RA Medina N., Meliado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portecelle D., Potwilk S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Serio S.J., Serio P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Wellenzegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RN Nature 390:249-256(1997).
RP CHARACTERIZATION.
RX MEDLINE=95318144; PubMed=7797557;
RA Perego M., Glaser P., Mantello A., Strauch M.A., Leopold K.,
RA Fischer W.;
RT "Incorporation of D-alanine into lipoteichoic acid and wall teichoic
RT acid in Bacillus subtilis. Identification of genes and regulation."
RJ J. Biol. Chem. 270:15598-15606(1995).
CC -I- FUNCTION: COULD BE INVOLVED IN THE TRANSPORT OF ACTIVATED
CC -I- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -I- PATHWAY: D-alanyl-lipoteichoic acid biosynthesis.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; X73124; CAAS1560.1; -
CC DR EMBL; Z99123; CAB15877.1; -
CC DR PIR; S39659; S39659.
CC DR Subtilist; BG10550; dlrb.
CC DR Pfam; PF03062; MBOAT1.1.
CC Transmembrane; Transport; Complete proteome.
CC FT TRANSMEM 7 POTENTIAL.
CC FT TRANSMEM 39 POTENTIAL.
CC FT TRANSMEM 80 POTENTIAL.
CC FT TRANSMEM 115 POTENTIAL.
CC FT TRANSMEM 144 POTENTIAL.
CC FT TRANSMEM 194 POTENTIAL.
CC FT TRANSMEM 238 POTENTIAL.
CC FT TRANSMEM 315 POTENTIAL.
CC FT TRANSMEM 366 POTENTIAL.
CC SQ SEQUENCE 395 AA; 46735 MW; 7A22C7B711FD6571 CRC64;
Query Match
Best Local Similarity 18.2%; Score 104.5; DB 1; Length 395;
Matches 85; Conservative 58; Mismatches 141; Indels 183; Gaps 22;
59 YHNYFYEYETVLIHFPHFTGISTAVFNGNLYSLCTIVQFILIR-LMGRTITAVL 115
4 YSSFLFLLGLLPTIILGNKRF---QAYNMFISIIILALFSDHLDG---VIAL 56
116 TFFCQMAVLAGY-YTATGNYD-----IKWTPH----- 145
57 CLFTWOUVLISGLYAKOKANSGFVFCGAVIASLPLFLSKIMPFSLHPQHHPPNLI 116
146 -----CVTLKILGLAVYFDGKQNSLSBQKRYAIRGVPSLLEVAQFSFYCAFLVG 200

```

```

Db 117 SFLGISVLTGKGVOLINEARDG-----LKEQ-----LP-LHRLLYFLFPPTISSG 162
Qy 201 POFSMNHMYKLVOG-----ELIDIPGKIPNSIIPALKRSLGLFY--LVGYTLIS-- 248
Db 163 P----IDRYRRVKBQAKMTKEEYADL-----LYGINKIFLFGFLPKFIIGAINITYF 212
Qy 249 ----PHITEDYLLTEDYDNHDFRCMYMLIWGKFLVLYKVTCLVTEGVCILTGL----- 300
Db 213 IMNLPALITHNKILG-----NLLYMGYSMYLFFDFGMYPMFANVGSYINGIKSPE 262
Qy 301 GFNFEEBKAKMKACANMKWLEPTNPRFTGTASF---NITNMAVARIYFKRLKFL 356
Db 263 NFN-----KPFISKNIKDFNRRWMSLSFWRDYVFMRFVFW 299
Qy 357 GNKE-----LSOGLSLF--LALWGHSGVLYVCFQMEFLIVYERQARLIQESP 405
Db 300 MTKKKMKKNMAVSNIGYFLFPLMGLGVHGLAPQYI----- 335
Qy 406 TLSKLAITVLQPFYLVQOTIHWLFMGYSMTAFCLFTWDKMLKLYK 452
Db 336 -----IYGLYHAVLMTGYNPF--EKMKKKYK 359

RESULT 15
CYB TINMA STANDARD; PRT; 379 AA.
AC 003553;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cytochrome b.
GN MTCYB OR COB OR CYTB.
OS Titamus major (Great titmouse).
OC Mitochondrion.
OC Bukariyoti; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Palaeognathae; Tinamiformes; Tinamidae; Tinamus.
OX NCBI_TaxID=30468;
RN [1]
RP SEQUENCE FROM N. A.
RA Lee K, Feinstein J, Cracraft J;
RT "Phylogenetic relationships of the ratite birds: resolving conflicts
RL between molecular and morphological data sets.";
RL (in) Mindell D.P. (eds.);
RL Avian molecular evolution and systematics, pp.1-1, Academic Press,
RL New York (1997).
CC -1- FUNCTION: Component of the ubiquinol-cytochrome c reductase
CC complex (complex III or cytochrome b-c1 complex), which is a
CC respiratory chain that generates an electrochemical potential
CC coupled to ATP synthesis (By similarity).
CC -1- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
CC b562) is low-potential and absorbs at about 562, and heme 2 (or BH
CC or b566) is high-potential and absorbs at about 566 (By
CC similarity).
CC -1- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
CC cytochrome c1 and the Rieske protein (By similarity).
CC -1- SIMILARITY: Belongs to the cytochrome b family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U76056; AAB6130.1;
CC InterPro: IPR005798; Cytb b6 C.
CC InterPro: IPR005797; Cytb b6 N.
CC Pfam: PF00003; cytochrome_b_C; 1.
CC Pfam: PF00003; cytochrome_b_N; 1.
CC PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
CC PROSITE: PS00193; CYTOCHROME_B_OO; 1.
CC Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW

```

```

KW Heme.
FT METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 183 183 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 197 197 IRON 2 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 379 AA; 42797 MW; 60B9D75CBEA430AA CRC64;

Query Match 4.0%; Score 104; DB 1; Length 379;
Best Local Similarity 21.1%; Pred. No. 1.4;
Matches 63; Conservative 41; Mismatches 114; Indels 80; Gaps 11;

Qy 46 ISIFLGYPALFYRHHLYFYKETYLIHLFHTFTGLSLAVFNGQLYHSLCTIVQLILR 105
Db 93 ICIVLHIGRGLYGSYL-KETWNIGIILLTLMTATFVG-----VLPMGQMS 140
Qy 106 LMGRTITAVLITRPFQNAVLLAGIYYATGNYDI-----KATMPHCUTLTKIG 154
Db 141 FWGAT--VITNLSAIPYIGQALVEWAWGFSVDNPTLTRFPALHFLPLIVLSLSIH 197
Qy 155 LAVDPDGGKDONSLSSQOKYAIRGVPSLLEVAGF-----SYFYGAFL----- 198
Db 198 LTFHESGSNNPLGIITQSDDKIPHPYFTTKDMGLFTLMPPLITLAFPPNPLGDEPNF 257
Qy 199 -----VGPOFSMNHMYKLVOGELIDIPGKIPNSIIPALKRSLGLFYLWGYTLISP-- 249
Db 258 TPANPLITPRHIKPEWYFLFAVAILRSIPNKLGG--VLALTASVLIF-----LSPL 308
Qy 250 HITEDYLLTEDYDNHDFRCMYMLIWGKFLVLYKVTCLVTEGVCILTGLGFGFEE 307
Db 309 HASKMRSLT-----FRPMGSLF-----MLIANLILITWIGSQPVED 346

```

Search completed: November 21, 2003, 14:15:12
 Job time : 19 secs

THIS PAGE BLANK (USPTO)

Db 292 -----GLFVI-----LYSIANALASISALAFS-----ID 315

QY 317 ANMKWLFETNPF-----TGTIASFNINTAMVARYFKRLKFL 356

Db 316 APKILSLDADPHFIKRLRINKRGTPTNGYMTILVSLIITPAFGIGNNELYKWL 375

QY 357 GN-KELSGISSLFLALMGHLSGYLVCFQMEFLIYVERQARLLIOESPTLSKLAITV 415

Db 376 LNLNSVVMFLRYLWFLAYLLNQHLKQADYMLV-----KKRLGLMGV----- 421

QY 416 LQPFYLVQOTIHLFMGYSMTAF-CLF-----TWDKRLKYKSYFLGHI 461

Db 422 -----RMCF-----FTAFACLLGMLPKRISYLANDPGTW--WFOV--GLNITTPVI 462

QY 462 FLSLFLPYI 472

Db 463 LNLGLILPYI 473

RESULT 2

US-08-669-656A-6

Sequence 6, Application US/08669656A

Patent No. 6451554

GENERAL INFORMATION:

APPLICANT: Wood, John N.

APPLICANT: Akopian, Armen N.

TITLE OF INVENTION: Ion Channel

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: ZENeca Pharmaceuticals

STREET: 1800 Concord Pike, P.O. Box 15437

CITY: Wilmington

STATE: Delaware

COUNTRY: USA

ZIP: 19850

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/669,656A

FILING DATE: 24-JUN-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Hohenschutz, Lisa D.

REGISTRATION NUMBER: 33,712

REFERENCE/DOCKET NUMBER: PHM.70086

TELECOMMUNICATION INFORMATION:

TELEPHONE: (302) 886-7466

INFORMATION FOR SEQ. ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 2132 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-669-656A-6

Query Match 4.1%; Score 105; DB 4; Length 2132;

Best Local Similarity 19.4%; Pred. No. 0.13;

Matches 113; Conservative 74; Mismatches 194; Indels 200; Gaps 28;

QY 7 GEGGTVALAGVYQSGFOE-----LSLNKLATSLGSEQLRL-----IISTF----- 49

Db 565 GEGGQIGVTGELTAAPRGPALDTTGQSFLSAGYLNFPFRQRAMSVASITSIEEL 624

QY 50 ----LGYFPAL--FYRHVLYFK-----ETYLHLFH-----TFGTLSIAVFNFGNQLYHSL 94

Db 625 EESKSKCPCLISFAKQYLIWECCKRKRKXKALFLVLVDPRFAELIT----- 672

QY 95 LCIIVLOFLRL--MGRITIAVLTT-----FCQMAVYLLAG--YYTATGNYOI 139

Db 673 LCIIVTVPMAMEHMYMTDAFDAMLQAGNIVFTVFTMEMAFKIIAFDPPYY-----FOK 727

QY 140 KWTMPHCYL-TKLIGLAVDPDGKDONSLSSEQOKAIRGVPSLLVAGSFYGAFL 198

Db 728 KNINIDCVIIVLSLELS-----ASKKGSLS-----VLKRLALDITGKSLFSLAGY 775

QY 199 VGPQFSNMHYKLVQ-----GELIDIPGKIIPNSII-----PALKRLSLGLFY 240

Db 776 LNEPFRQRAMSVASIMTSVIELESKLCPCPLISFAQKYLWECCKRKRKXKALFE 835

QY 241 LVGYTLSPHIT-----EDYLLTEYDHNPFMRFRMVLINCKFLYKXVTCW 288

Db 836 LVTDFPABLITTLCTIVNTVFNAMEHYPMTDADF-----ANLQAGNIVFTVFTWE 886

QY 289 LVTEGVCLITGLGFNGFPEEKAKMDA--CANMKWLFETNPRFTGIASF----- 337

Db 887 MAFK--IIPDPYVYFQK--KNINIDCVIIVTSLLELSASKKGSLSVLKSLRLAVF 939

QY 338 -----NINTNAMVARYI FKRLKFLGNKELSGSLFLAL----- 372

Db 940 KLAKSMPILTNT--LIKIIINSVGLGNLTFLIATVIFALVGLKQLSEDYCRKDGVS 996

QY 373 -WHGLHSGYLVCFQMEFLIYVERQARLLIOESPTLSKLAITVLOPFYLVQOTIH-- 428

Db 997 VVNGSKRLRWMCDFHSEFLV-----FRLLGGEIENM 1029

QY 429 WLFMGYSMTAFCLFTWDMKWLKYKSYFLGHIFFLSLFL 469

Db 1030 WYCMVEVSQKSLITL-----LFLTVMVLGNLVANLFLAL 1063

RESULT 3

US-09-134-001C-4885

Sequence 4885, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4885

LENGTH: 412

TYPE: PR

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4885

Query Match 4.0%; Score 102.5; DB 4; Length 412;

Best Local Similarity 16.0%; Pred. No. 0.02;

Matches 76; Conservative 74; Mismatches 165; Indels 159; Gaps 17;

QY 73 FHTFTGLSIAYFNFGNQLYHSLCTVLOFLIRLMGR-----TTVAULTPCFQWA 123

Db 12 YCTFTFPFLAFL-----VLIPYILIGFLGKSKSYIYNGISTAMIMYIIPASDKH 59

QY 124 YLLAGYVYATANGNDIKWTMPHCYTLTKLIGLAVDPDGKDONSLSSEQOKYAIRGVPS 183

Db 60 NLLGQKYLDAQ-----LFCRIIVYQVNLIMVYYSRQKKNFT--KFVITMVL 108

QY 184 LLEVAGSFYGAFLVGPQFSNMHYKLVQGLIDIPGKIIPNSIIP-----A 230

Db 109 ILPLAIVKILQSSWLG-----HOIHFHESKLIIEFVFLGISYVTFKSVQVIMEIRDS 162

QY 231 LKRLSLGLFYLVGYTLSPHIT-----EDYLLTEYDHNPFMRFRMVLINCK 278

Db 163 IKEIVG--KLQFISFPFTISSGPIDRYKRFVKDDKVPVSGAEYRELVAKAIHMIWG- 219

QY 279 FVLKYKVCVWLVTBVCILITGLGFNGFPEEKAKMDACANMKWMLF----- 324

Db 220 -FLYKIIIVLYIQTAVNPLQMLHGFTH-----MPLMYAVSYLFFDFAGYSLFAIVS 274
QY 325 -----ETNPRFTGTISFNINT--NMMVARYIF-----KRLKFLGNKELSGQ----- 364
Db 275 YLYGKITPPNFKOPFYAKNKKDKWNRMHMTLSFWPDCIYMRSLFYMSRKKLSQFAMS 334
QY 365 -----LSLFLALMHGHSGLVCFQWEPLIVEROAARLQESPTLSKLAITVLOPF 419
Db 335 NNAVYFLNPFIMGVHGLE-----V 353
QY 420 YIVVQOTIH-WLPMGYSMTAFCLFTWDKMLKVYSIYFLCHIFLSLFLPI 472
Db 354 YIVVGLYHVAALFYGYY-----YERMRKQHPRMQNGFTVAISIIIFHFV 400

RESULT 4
US-09-328-352-7069
Sequence 7069, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 7069
LENGTH: 443
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-7069

Query Match
Best Local Similarity 21.3%; Score 99.5; DB 4; Length 443;
Matches 61; Conservative 42; Mismatches 90; Indels 93; Gaps 11;

QY 11 TVVALA--GVLOSGF-----OELSNKLATSGASEQA--LRLLISIFAGYFEL 56
Db 185 TVVALACIGLFRFLIPQVAAAPDDIDMNFIRTSQTIKYSKLPVLTILGNSW-- 242
QY 57 FYRHYLFYKETYL-----IHLEHPTGSLAVFNFGNQLYHSLCIVLQPLRL 106
Db 243 ---YMFYGATYLTQIPQLTQONLHASENVVSLTLTFVSGIV--GSLCRKIGSGEINI 297
QY 107 MGRITIAV-LTTCFQMAVLA-----GYTTATGNYDIKMTMPHCVLTLLKIGLAVY 159
Db 298 KWPVIGAVGLTVALYLAAGLAVPERTGALLTLKDFQGWYVYHMAVLTLLGIS-- 354
QY 160 FDGKDONSLSBQKAYAIRGVPSLLEVAGFSYFYGAFVGGPOFSM-----NHWKLV 212
Db 355 -----GGFYIVPLXAMQAVSPSHRAV 379
QY 213 QGELIDIPKIPISIIIPALRSLGLFVYGVYTLSPHTEYLLT 258
Db 380 AA-----NNILNAVFMVSSAIFSLILSVLKIDIKILFIIT 415

RESULT 5
US-09-352-991A-29870
Sequence 29870, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 29870
LENGTH: 243
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29870

Query Match
Best Local Similarity 23.3%; Score 98.5; DB 4; Length 243;
Matches 65; Conservative 36; Mismatches 91; Indels 87; Gaps 13;

QY 225 NSIIPALKRLSLGLFVGYTLSP--HITEDYLLTEDYDNNHFWFCMYMLWKGCV 280
Db 28 DGLVTLTKLLSL--LIGLLAVPLALMRVSKQPLV-----NFPAM 67
QY 281 LYKYTCMLVTEGVCILTGLGRNGPEKAKWDACANNKWLFEINPRFTGTISFNIN 340
Db 68 LVTY-----VIRGTPLVQLFLIY--GLAOFDAVESALMPWLSNAPCACL-AFAIN 118
QY 341 TNMVARVYFKRLKFLGNKEL-----SQGLSLFLALMHGHSGLVCFQWEPLIVERO 396
Db 119 TSVYTBELIAGSLKATPHGEIEAKAMGMSRL-----KMYRILLPSAL 162
QY 397 AARLQESPTLSKLAITVLOPEYVVOQTIHMLFMGYSMTAFCLFTWDKMLKVYSIYF 456
Db 163 RRLPQYSNEVIMMQDTSLASIVTLVDIT-----GAARTV-----SOY 203
QY 457 L-----GHIFLSLFLPIYHKA-----NWPBK 480
Db 204 LPFEAFITAGLFYLCFTLVRLFLKLABRRMLAYLAPRK 242

RESULT 6
US-09-107-532A-3878
Sequence 3878, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A. Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
City: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Daneke
REGISTRATION NUMBER: 40,469
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3878:
SEQUENCE CHARACTERISTICS:
LENGTH: 401 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...401
SEQUENCE DESCRIPTION: SEQ ID NO: 3878
US-09-107-532A-3878

Query Match 3.8%; Score 98.5; DB 4; Length 401;
Best Local Similarity 18.6%; Pred. No. 0.052;
Matches 75; Conservative 45; Mismatches 136; Indels 147; Gaps 17;

QY 60 HYLKYEKYLHLHFTFTGLSIAYFNGNQLYHSLICVLQPLI---LRLMGRITIT--- 112
DB 5 HMLPYAEPPYFLL-----LAIA-----LVPILVSLIMKRLPYQSAVTLTF 47
QY 113 -----AVLTFCFQMAVLLAGYYTATGNNDIKMTMPCVLTGLGL- AV 157
DB 48 LFVSPGNGVMOGAGALLGYVLMQTLVNGYFVYRQRKQGTGMPRAVTLAIPLFTVT 107
QY 158 DYFDGKDQNSLSBQKVAIRGVPSLIEVAGFSY----- 192
DB 108 PFEHKG-----ASLIGLIGISYLTFSYQVIMELRDGMIKTKYKPRY 150
QY 193 -----FYCAFVLP-----QPSMNHMYMLVQGLIDIDPK-IENSIILAKRLSLGLFYL 241
DB 151 IQPLLFPTLTSSGPDYRRREKQDHPVDRKVDLGGKVONIFVGLYKFTIG--YY 208
QY 242 VGYTLSPHITEDYLLTEDYDNP--FWFRGMYMLMGKFLYKVCMLTEGVCLITG 299
DB 209 FGVYLL- PVERIAL-----QHGISALVGMVYVSMYLFDPBAGSLAVGTSYMG 261
QY 300 LGFNGFEKGAKKADCANMKVWLFEINPRTGTIASFNI-----NTNAVARYIF 350
DB 262 -----YQTPINFNKPLISVNIKEFNMWMHTLSFWRDYYV 297
QY 351 KRL-----KELNGKELSGLS-----LLFLALMHGHSGLV 382
DB 298 MRLMFPMMKKKPFKSKIVMSNIGYFALFLMGLIMGLTWYIV 340

RESULT 7

US-09-328-352-8050
Sequence 8050, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328.352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 8050
LENGTH: 488
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-8050

Query Match 3.8%; Score 97.5; DB 4; Length 488;
Best Local Similarity 23.4%; Pred. No. 0.089;
Matches 55; Conservative 32; Mismatches 101; Indels 47; Gaps 9;

QY 11 TVALAGVLOSGFELSLINKLATSIGASEQALRLIISIFLGYPPALFYRHLYFYKETYLI 70
DB 98 SLVGAALVLIIVAAIHNHQLTALFAQNEALTL-----AAGVGHAMSK-----V 142
QY 71 HLFHTFTGLSIAYFNGNQLYHSLICVLQPLIIRLMGRITITAV---LTTFCFQMAVLL 126
DB 143 HLFELFVCGCFVGAITFTASV-----AYGTLAAKKMAKTTISGAMVVKVQALLF-VAMLA 195

QY 127 AGYYTATGNVDIKMTMPCVLTGLGLAVDPDGKDQNSLSBQKVAIRGVPSLLE 186
DB 196 CGFTFTTGNTAFWANTALALAFGWVWIP--VGGD-----MPVVSLL- 239
QY 187 VAGSFYFAGFLVGPQPSMNHMYMLVQGLIDIPGKIPNSII-PALKRLSLGLFY 240
DB 240 -----NSISGMAAGIGFTLENMMLIVAGSLVSGGALLSYIMCAMNRSIINVL 290

RESULT 8

US-09-252-991A-32219
Sequence 32219, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 32219
LENGTH: 545
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32219

Query Match 3.7%; Score 96.5; DB 4; Length 545;
Best Local Similarity 21.5%; Pred. No. 0.14;
Matches 62; Conservative 43; Mismatches 105; Indels 79; Gaps 18;

QY 221 GKIPNSIIPALKRLSLGLFYLVGYT-----LSPHITEDYLLTE-DYDNPFW 267
DB 125 GRINAF--AVGRRRMGWMLLVFPATTIANYIDRAALGVMPILAEKSWTAMDYANINFW 182
QY 268 FRCMY--MLMGKV-----LYKYVTCMLTBSV--CILTGLGFN-----GFEKK 308
DB 183 FQVGYAIGFILOGFIDKVGKRAFFLAVLWLSIATGAHLGATSAGAFMVCRTFGLTE- 241
QY 309 GKAKMDACA-NMKVWLETPNPTGTIASFNINMTNAVARYIFRLKFLNGKELSGLSL 367
DB 242 -AANYPACVKTTRLM-PPAGERAVAT-GIFNAGTNV-----GAVTPALLP 284
QY 368 LFLALMHGHSGLVLCFQMEFLVI-----VERQARLI--QESPTLSKLA 411
DB 285 LILSVW-GWQAAFLCNSALGLVWLVPFWRMYVNPDEHPRVKQSELEYIQQDEPPATRV- 342
QY 412 AITVLOPPYLVQQTIIHMLF-MGYSMTAFCLFTWDRKMLKY-KSIYFLG 458
DB 343 -----PFSQILRRRGTAFAFALAYSIAPVFWFYLWLPFLNQOYGLG 385

RESULT 9

US-09-328-352-7176
Sequence 7176, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328.352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 7176
LENGTH: 551
TYPE: PRT
ORGANISM: Acinetobacter baumannii

US-09-328-352-7176

Query Match 3.7%; Score 95; DB 4; Length 551;
Best Local Similarity 22.3%; Pred. No. 0.2;
Matches 55; Conservative 38; Mismatches 96; Indels 58; Gaps 11;

QY 257 LTEDYDNHFP-----WFCMVMILGWKFLVLYKVTGWL-----VTEGVCIITGLGPN 303
DB 53 LTPRIDGPFVHKPPLHMLSMFMELFG-----VHVWVLRVPLVAGTLMVLGL--- 102
QY 304 GFEEKGRKMDACANMKWLEETNPFVGTIASFV--INTNAVARIYFKLKFQNKEL 361
DB 103 -FLPVKHISSNVAGLTIVILATMLLFFGSSQYINHDLLTSMITISVLCFVDF---TI 157
QY 362 SQGLSLFLALMHGLHS-----GYLVCFQWEFLVIVEROARLIQESPTLSLAIT 414
DB 158 SARSKIPLGLYAGAAFLSKGLIGILIPGMILPMLYTKOMKKIPSLPLMLLPL 217
QY 415 VLQPFYLVQ---QTIHMLFMGYSMTAFCEFTWDK--WLKVYSIYFLGHIFPLSLF 467
DB 218 IVSPMLYVQSKVPQFLHYFFIDQFNPFSSKEFNKKQPM-----CFYLMILF 265
QY 468 I--LPYI 472
DB 266 VSFLPWL 272

RESULT 10

US-09-328-352-7092
Sequence 7092, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328.352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 7092
LENGTH: 308
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-7092

Query Match 3.6%; Score 93; DB 4; Length 308;
Best Local Similarity 20.4%; Pred. No. 0.14;
Matches 70; Conservative 49; Mismatches 114; Indels 110; Gaps 15;

QY 11 TVVALAGVLOSQFQELSLINKLATSIGA-----SEQALRLIISIFIGYPALFYRYL 62
DB 12 TLIGLSALIMASWGVFKHITTAIGPDVGITLISLNLILLI-IFVVPFKLSKXYL 70
QY 63 F-----YKETYLIHLFHTFTGLSIAYFNGQLYHSLICIVLQPLIRLMGRTITAV 114
DB 71 ILGAILFLAYELCFSPALAYSKTAQOALEVSIYVNLMPSL--TVLAFVIFRELKENVPLI 128
QY 115 LTTECFQAYLLAGYYTATGNYD-----IKTMTPCVUTL 150
DB 129 L-----GLLISIGIIFQTGNGDPSLGRVVDNFHNSPLSYLAEIGAIMAF-YCVILTK 182
QY 151 KLIGLAVDYFDGQKQNSLSSEQKYAIRGVPSLLEVAGSYFAGFLVQPGFSNNHYMK 210
DB 183 KM-----GKQONPIS-----IFFLGV-----ALTLMK 205
QY 211 LVQGLIDIPGKIPIISIPAL-KRLSLGLFY-----LVGYTLSPHTE-- 253
DB 206 LIFSQVSLPSLDLNTIILSLIVASAIGLYAAMNIGIIGHNITMLVVASYFTPIISSL 265
QY 254 DYLLTEDYDNHFPFRCMMLIWGFVLAKYVTCVLVBEVCV 296
DB 266 AMLVQLTELISLFW-QGTAMVYTAGSLI-----CWLSTNMAVI 301

RESULT 11

US-09-066-047-7
Sequence 7, Application US/09066047A
Patent No. 6306394
GENERAL INFORMATION:

APPLICANT: MURPHY, Cheryl
STOREY, James
BELTZ, Gerald A.
COUGHLIN, Richard T.
TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND METHODS OF
USE OF GRANDLOCYTIC ERLICHIA

NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: HALE AND DORR LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: United States
ZIP: 02109
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,047A
FILING DATE: 24-Apr-1998
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/044,869
FILING DATE: 25-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Superko, Colleen
REGISTRATION NUMBER: 39,850
REFERENCE/DOCKET NUMBER: 106,941.156
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 526-6000
TELEFAX: (617) 526-5000
INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 506 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-066-047-7

Query Match 3.6%; Score 93; DB 4; Length 506;
Best Local Similarity 20.7%; Pred. No. 0.29;
Matches 75; Conservative 45; Mismatches 133; Indels 110; Gaps 15;

QY 27 SLINKLATSIGASBOALRLIISIFIGYPALFYRYLYLHFTFTGLSIAYPNF 86
DB 164 SFGMLVSLPFTTKVALYVLKFFSGESITLY-----FCIFSIYAIRAF 208
QY 87 GNQLYHSLICIV-----LQFLIRLMGRTITAVLTTCFQMAV-LLAGYYTA---TG 135
DB 209 LEQVVRRLMAYMFGQAGLMMAGCPISPSDLIIIVQLSFSVLVOLLLGMFADSVVRSKG 268
QY 136 NYDILKMTMPHCVLTLKLIGLAVDYFDGKQDQNSLSSEQKYAIRGVPSLLEVAGSYFYG 195
DB 269 HVDIN-RMAGC--FKLASME-----AMGCIYALMLNIGCFPMTAG 304
QY 196 AFLVGPQFSNN-----HYMKLVQGLIDIPGKIPIISIPAL-KRLSLGLFYLVGYTL 246
DB 305 FVTGMLHMLWLOSFDVWLKLYMQPMLGWL-----FASNGMLPWL---AC 348
QY 247 LSPHTEYLLTEDYDNHFPFRCMMLIWGFVLAKYVTCVLVBEVCVITGLGPNFE 306
DB 349 LKPCST-----TPEVAPSPFSSKLSIIML-----SLITVSGVLVYEGGLIFS--E 391

QY 307 EXGAKMADACANMKWLFETNPRFTGTASFNINTNAVARYIPKRLKELGNKELSGOGLS 366
DB 392 HKFVTFPGVATKTLWL-----CGVYLF-----FILPRRQPLGNYESAIGDS 433
QY 367 LIF 369
DB 434 WYV 436

RESULT 12

US-09-107-532A-5141
Sequence 5141, Application US/09107532A
Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: LYNN A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Arinello, Pamela Dencke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 5141:

SEQUENCE CHARACTERISTICS:

LENGTH: 793 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...793

SEQUENCE DESCRIPTION: SEQ ID NO: 5141:

US-09-107-532A-5141

Query Match 3.6%; Score 93; DB 4; Length 793;
Best Local Similarity 20.8%; Pred. No. 0.57;
Matches 64; Conservative 47; Mismatches 115; Indels 82; Gaps 13;

QY 12 VVALAGVQSGFQELSLKATSLGASDAQALRIISIFLGYFALFYRYLYFYKETYLIH 71
DB 335 IIVLSAVYFAGIMVLLSTI-TNLSAISRLFOVL-----PFS-----FNLDDQTVNL 382
QY 72 LFTFTGL-----SIAYFNG-----NOLYHSLGLVQLFLRLRMGRT- 110
DB 303 VGFLLLGLARLALSIVYKKAFLPTIALSLGILNITRRLISWQLIIVYVILILAVWSKRE 442
QY 111 -----ITAVLITPFCFQWAVYLLAGYY-----TATGNY-----DIKWTMP 144

DB 443 FYRTKFWYVSWGAILIFDASLFGFLF-IYVIALGYHSGRFLVGRITGNHFLFSSDVMPSG 501
QY 145 HCVLTKLIGLVANDYFDGSKDNLSSEQOK-----YAIRVPSLLEVAGSYFY- 194
DB 502 LIGLISLIALVALYQVLAEDTYYTKLGSSEYKDFRDLLETGYGTATSHQLRLSGSYFY 561
QY 195 -----GAFVVGQFQSMNHMYKLVQVELIDIPKIPNSIIPALKRSLGLFVLYGTYLLSP 249
DB 562 QEDDDQVVFQYQIRGNKC--FVLGDYVGNPRKMAATNQPFKQAD-----CLGQLAFY 614
QY 250 HITEDYLL 257
DB 615 KISESYV 622

RESULT 13

US-09-097-889-25
Sequence 25, Application US/09097889
Patent No. 6218117

GENERAL INFORMATION:

APPLICANT: Herinstdt, Corrina
APPLICANT: Ghosh, Soumitra S.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE
TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA RATIOS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/097,889

FILING DATE: 15-JUN-1998

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Roseman Ph.D., Stephen J.

REGISTRATION NUMBER: 43,058

REFERENCE/DOCKET NUMBER: 660088.417

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 380 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-09-097-889-25

Query Match 3.6%; Score 92; DB 3; Length 380;
Best Local Similarity 20.1%; Pred. No. 0.24;
Matches 54; Conservative 32; Mismatches 75; Indels 108; Gaps 16;

QY 201 PQFSNMHMYKLVQVELIDIPKIPNSIIP-----ALKRSLGLFVLYGTYLLSP 249
DB 3 FMRKINPLMKLINSFIDL-TPSNISAMWNFSLGACILITTTGLFLAMHY----SP 57
QY 250 -----HITED-----YLTEDYDNHPFWRQVM-----LWQKFWLYKXYT 286
DB 58 DASTRFSSIAHITRDVNYGWIIRYL--HANGASNFICLPLHIGRGLYGSFL--YSE 111
QY 287 CMLYTEGVCIL---TGLQFNGFEEKXKAKMADACANMKWLFETNPRFTGTASFNINTNA 343

Db 112 TWNI--GIILLATMATAFMGY-----VLPW---GOMSEW-----GA 143
Qy 344 WVAIYIFRKLKFNKLSQGLSLFLALMGLHSGYLVCFOMEFLIVYERQARLIOE 403
Db 144 TVTNLSAIPYIGTD-----LVQW---IWGYSV-----D 171
Qy 404 SPTLSKLAATVLOPFYLVQOTIHWLFM 432
Db 172 SPTLRPFTHFLPFIATLATHLFL 200

RESULT 14
US-09-098-079-25
Sequence 25, Application US/09098079
Patent No. 6489095
GENERAL INFORMATION:
APPLICANT: Herrnstadt, Corrina
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Cleveland, William
APPLICANT: Fahy, Boia F.
APPLICANT: Davis, Robert E.
TITLE OF INVENTION: DIAGNOSTIC METHOD BASED ON QUANTIFICATION OF
NUMBER OF SEQUENCES: 26
EXTRAMITOCHONDRIAL DNA
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098, 079
FILING DATE: 15-JUN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman Ph.D., Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 660088, 416
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-6031
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 380 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-098-079-25

Query Match 3.6%; Score 92; DB 4; Length 380;
Best Local Similarity 20.1%; Pred. No. 0.24;
Matches 54; Conservative 32; Mismatches 75; Indels 108; Gaps 16;

Qy 201 PQSMNHMKLVQGLIDIPKIPNSIIP-----ALKRLSLGLFYLVTLLSP 249
Db 3 PMRKINPLMKLINHSIDL--TPSNISAMWNGSLGACILQITGTLFLAMHY---SP 57
Qy 250 -----HITED-----YLTEDYDNHPFRCMTM-----LIMGKFLVLYKYVT 286
Db 58 DASTAFSSIAHTIDVNYGWIIRYL--HANGASMFICLFLHIGRGLYGSFL---YSE 111
Qy 287 CMLTEGVCI--TGAGFNGFEKGRKAKMDACAMKVMLEFETNPRFTGTIASFNINTNA 343
Db 112 TWNI--GIILLATMATAFMGY-----VLPW---GOMSEW-----GA 143
Qy 344 WVAIYIFRKLKFNKLSQGLSLFLALMGLHSGYLVCFOMEFLIVYERQARLIOE 403

Db 144 TVTNLSAIPYIGTD-----LVQW---IWGYSV-----D 171
Qy 404 SPTLSKLAATVLOPFYLVQOTIHWLFM 432
Db 172 SPTLRPFTHFLPFIATLATHLFL 200

RESULT 15
US-09-328-352-5397
Sequence 5397, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328, 352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 5397
LENGTH: 470
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-5397

Query Match 3.6%; Score 92; DB 4; Length 470;
Best Local Similarity 21.2%; Pred. No. 0.33;
Matches 66; Conservative 46; Mismatches 124; Indels 76; Gaps 16;

Qy 138 DIKWTM--PHCVLTUKLIGLAVDFDGGKQNSISEQOKYAIRGVPSLL--EVAGSFYFY 194
Db 73 DTSWILSTALVLMITPIGLALFYGGMVRKKNVLTWMPSLAAILVSLMLWVINGSIAP 132
Qy 195 GAVLVERQFSMNHMKLVQGLID-IPKIPNSIIPALKRLSLGLFYLVTLLSPHTE 253
Db 133 SG--TGAYFG--DISKAMLVNVAFDALSGTIPESLFIQW---TFALITVALISGSA- 184
Qy 254 DYLLTEDYDNHPFRCMTMIMGKFLVLYKYVTCMLTTEGVCIITLGAFNGFEKGRAKW 313
Db 185 -----DMKKSAPFAFAIWM-VLVVAPITMV-----W 212
Qy 314 DACAMKVMLEFETNPR--FTGTIASFNINTNAVARIYIFRKLKFNKLSQ----- 363
Db 213 AADG---WLFKAGALDPAGSTVHVHNSGVAALVAAVMLKRLIGLGRSMAPHNLTTLVI 268
Qy 364 GLSLFLALMGLHSGYL-----CFQMEFLIVYERQAA-----RLIQESTL----- 407
Db 269 GASLWVG-WFGFNGSALGAGARASMAILVTQVAAAAAASFMLVVERMIRGKASVIGGA 327
Qy 408 -SKLAATVLOP 418
Db 328 SGAVAGLVITP 339

Search completed: November 21, 2003, 14:17:10
Job time : 22 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: November 21, 2003, 14:16:08 ; Search time 35 Seconds

(without alignments)
2540.185 Million cell updates/sec

Title: US-09-938-803-6

Perfect score: 2580

Sequence: 1 MASSAEGDEGTVALAGVLQ.....ILPYIKAMPKREKXKME 487

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2580	100.0	487	9	US-09-938-803-6
2	350	13.6	488	12	US-10-053-248-6
3	310	12.0	57	12	US-10-029-386-29742
4	231	9.0	425	12	US-09-821-812-9
5	215	8.3	58	9	US-09-864-761-33415
6	150	4.8	342	12	US-10-205-219-192
7	123.5	4.8	141	12	US-09-821-812-7
8	111	4.3	392	12	US-10-287-274-375
9	105	4.1	2132	12	US-10-202-824-6
10	101	3.9	613	9	US-09-815-242-13958
11	99.5	3.9	263	11	US-09-769-787-119
12	99.5	3.9	663	12	US-09-781-880-5
13	96.5	3.7	587	9	US-09-815-242-5694
14	96.5	3.7	603	9	US-09-815-242-12667
15	95.5	3.7	286	10	US-09-731-872-287

16	95.5	3.7	286	12	US-09-876-997-287	Sequence 287, App
17	93.5	3.6	613	9	US-09-815-242-10225	Sequence 10225, A
18	93.5	3.6	697	10	US-09-881-752A-76	Sequence 76, App1
19	93	3.6	819	10	US-09-796-720B-4	Sequence 4, App1
20	93	3.6	819	12	US-10-146-733-24	Sequence 25, App1
21	92.5	3.6	316	11	US-09-510-332-121	Sequence 121, App
22	92.5	3.6	380	9	US-09-098-079-25	Sequence 2, App1
23	91.5	3.5	473	15	US-10-000-151B-2	Sequence 12, App1
24	91	3.5	570	11	US-09-922-225A-12	Sequence 8, App1
25	91	3.5	627	15	US-10-090-455-8	Sequence 2, App1
26	91	3.5	646	12	US-10-079-087-2	Sequence 2, App1
27	91	3.5	646	14	US-10-154-452-4	Sequence 9, App1
28	91	3.5	646	14	US-10-154-452-4	Sequence 4, App1
29	91	3.5	646	14	US-10-154-452-8	Sequence 8, App1
30	91	3.5	646	15	US-10-090-455-2	Sequence 2, App1
31	90.5	3.5	403	9	US-09-738-626-3904	Sequence 3904, Ap
32	90.5	3.5	622	9	US-08-843-558-11	Sequence 11, App1
33	90.5	3.5	637	8	US-08-879-337-8	Sequence 8, App1
34	90.5	3.5	1564	10	US-09-801-368-244	Sequence 244, App
35	89.5	3.5	338	10	US-09-976-605-6	Sequence 6, App1
36	89	3.4	388	9	US-09-741-669-370	Sequence 370, App
37	89	3.4	388	9	US-09-912-020-378	Sequence 378, App
38	89	3.4	874	10	US-09-963-766-6	Sequence 6, App1
39	88.5	3.4	206	11	US-09-791-279-129	Sequence 129, App
40	88	3.4	305	11	US-09-510-332-135	Sequence 135, App
41	88	3.4	664	10	US-09-898-533-2	Sequence 2, App1
42	88	3.4	664	15	US-10-176-847-40	Sequence 40, App1
43	88	3.4	697	10	US-09-989-919-108	Sequence 108, App
44	88	3.4	697	11	US-09-919-039-200	Sequence 200, App
45	88	3.4	697	12	US-10-024-298A-77	Sequence 77, App1

ALIGNMENTS

```
RESULT 1
US-09-938-803-6
; Sequence 6, Application US/09938803
; Patent No. US20020076762A1
GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Tang, Y. Tom
; APPLICANT: Lal, Preeti
; APPLICANT: Reddy, Roopa
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Yang, Junning
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: FULL-LENGTH EXPRESSED GENETIC MARKERS
; FILE REFERENCE: PF-0695 US
; CURRENT APPLICATION NUMBER: US/09/938,803
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/311,894
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 6
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte Clone 1928920
US-09-938-803-6

Query Match      100.0%; Score 2580; DB 9; Length 487;
Best Local Similarity 100.0%; Pred. No. 8.9e-239;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MASSAEGDEGTVALAGVLSGFOELSTNKATLSGSEOLRLITISFTLCYPPALFVRH 60
DB      1 MASSAEGDEGTVALAGVLSGFOELSTNKATLSGSEOLRLITISFTLCYPPALFVRH 60
QY      61 YLFYKETYLIHLFTFTGLSIAYFNFGNLVHSLICIVLQFLIRLIMGRITAVLTTPCF 120
```

```
Db 61 YLFYKSTYIHLFHTFTGLSIAYFNFGNOLYSLLCTIVAQFLRLMGRTITAVLTTCF 120
Qy 121 QMAVLLAGYYTATGYNDIKMTMPHCVLTKLIGLVDPDGGKDNSSSQOKXAIAG 180
Db 121 QMAVLLAGYYTATGYNDIKMTMPHCVLTKLIGLVDPDGGKDNSSSQOKXAIAG 180
Qy 181 VPSLLEVAGSFYFGAFLVGPQPSNMNMYKLVOGELIDIPGKI PMSIIPALKRLSLGLFY 240
Db 181 VPSLLEVAGSFYFGAFLVGPQPSNMNMYKLVOGELIDIPGKI PMSIIPALKRLSLGLFY 240
Qy 241 LVGYTLISPHITEDYLLTDDYDNPFPFQMTMLNGKVLKYVTGCVLTGVCILTGL 300
Db 241 LVGYTLISPHITEDYLLTDDYDNPFPFQMTMLNGKVLKYVTGCVLTGVCILTGL 300
Qy 301 GENGFEKAKAMDACANNKMWLFETNPRFTGTIASFNNTAMVARYFKRLKPLGNKE 360
Db 301 GENGFEKAKAMDACANNKMWLFETNPRFTGTIASFNNTAMVARYFKRLKPLGNKE 360
Qy 361 LSGSLSLFLALMHGSLGYLVCQMEFLIVEROARLIDESPPLSKLAAITVLQPPY 420
Db 361 LSGSLSLFLALMHGSLGYLVCQMEFLIVEROARLIDESPPLSKLAAITVLQPPY 420
Qy 421 YLVQOTIHWLFNGYSMTAFCLFTWDMKLYKYSIYFLGHI FPLSLFILPYIHKAMPK 480
Db 421 YLVQOTIHWLFNGYSMTAFCLFTWDMKLYKYSIYFLGHI FPLSLFILPYIHKAMPK 480
Qy 481 EKLKME 487
Db 481 EKLKME 487
```

```
RESULT 2
US-10-053-248-6
; Sequence 6, Application US/10053248
; Publication No. US20030144188A1
; GENERAL INFORMATION:
; APPLICANT: Lin, Biaoyang
; TITLE OF INVENTION: Androgen Regulated Nucleic Acid
; FILE REFERENCE: P-IS 4814
; CURRENT APPLICATION NUMBER: US/10/053,248
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-053-248-6
```

```
Query Match 13.6%; Score 350; DB 12; Length 488;
Best Local Similarity 24.4%; Pred. No. 6,3e-25;
Matches 108; Conservative 82; Mismatches 177; Indels 76; Gaps 13;

Qy 55 ALFPRHYLFYKST--YLHLFHTFTGLSIAYFNFGNOLYSLLCTIVQFLRLMGRTIT 112
Db 39 ALMPFRYLSHSSKTSFIRHVAVTLGLYLALFCFGYALHFLVQSGISYICMIITGVE-- 96
Qy 113 AVLTPFCFQMAVLLAGYYTATGYNDIKMTMPHCVLTKLIGLVDPDGGKDNSSSQOK 170
Db 97 -----NMNPMIITOKITSLACIHDGMRKDE-ELT 128
Qy 171 SROOKAIRGVPSLLEVAGSFYFGAFLVGPQPSNMNMYKLVOGELIDIP-----GK--- 222
Db 128 SROOKAIRGVPSLLEVAGSFYFGAFLVGPQPSNMNMYKLVOGELIDIP-----GK--- 222
Qy 223 -----IPNSIIPALKRLSLGLFYLVGYTLISPHITEDYLLTDDYDNPFPFQMTML 274
Db 189 QYERTPEPSPNTAV-VQKLVQSGISLFLHITICT-LPVSYNIDENHQAASPTKIYLY 246
Qy 275 IYKGFVLYKYVTCWLVTEGVCILTGLGENGFEKAKAMDACANNKMWLFETNPRFTGT 334
Db 275 IYKGFVLYKYVTCWLVTEGVCILTGLGENGFEKAKAMDACANNKMWLFETNPRFTGT 334
```

```
Db 247 ISLARPXYFAWTLADAINNAGFGRGYDENGARWDLISNLRIQIEMSTSEKML 306
Qy 335 ASFNNTAMVARYFKRLKPLGNKELSQSLFLALMHGSLGYLVCQMEFLIVEROARL 394
Db 307 DWNMIQTALMKRVCEYRTSPRTQ-----TFILSAIMHGVYPCYIYTLFGLVMTL-- 359
Qy 395 ROARLIDESPPLSKLAAITVLQPPYLVQOTIHWLFNGYSMTAFCLFTWDMK 447
Db 360 --AARAMNPNFHYRTESQKL-----FYDVTIWTIVQVAISTYVFPVLLSIKS 409
Qy 448 LKVKYSIYFLGHI FPLSLFILP 470
Db 410 LTFYSWYCCILIGLIVLLELP 432
```

```
RESULT 3
US-10-029-386-29742
; Sequence 29742, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 29742
; LENGTH: 57
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR12.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.5
; OTHER INFORMATION: SWISSPROT HIT: P54336, EVALUATE 8.00e+00
US-10-029-386-29742
```

```
Query Match 12.0%; Score 310; DB 12; Length 57;
Best Local Similarity 100.0%; Pred. No. 2,5e-22;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 292 EGVCLTGLGNGFEKAKAMDACANNKMWLFETNPRFTGTIASFNNTAMVARY 347
Db 2 EGVCLTGLGNGFEKAKAMDACANNKMWLFETNPRFTGTIASFNNTAMVARY 57
```

```
RESULT 4
US-09-821-812-9
; Sequence 9, Application US/09821812
; Publication No. US20030166520A1
; GENERAL INFORMATION:
; APPLICANT: Lin, Biaoyang
; TITLE OF INVENTION: Androgen Regulated Prostate Specific
; FILE REFERENCE: P-IS 4373
; CURRENT APPLICATION NUMBER: US/09/821,812
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Homo sapiens
```


FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(425)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-821-812-9

Query Match 9.0%; Score 231; DB 12; Length 425;
Best Local Similarity 25.2%; Pred. No. 1.3e-13;
Matches 69; Conservative 45; Mismatches 122; Indels 38; Gaps 7;

QY 118 FCFQWAVL-----LAGYYTAANGNYDKMTMPHCVLTLKLGIAVDYFG--GKQNSLSS 171
DB 11 FVFALGYTLVQCVRVYLFYDGYSDPSGPMWITTKITSLACIHGMFRKE-ELTS 69
QY 172 EQQKAIAGVPSLLEVAGFSYFYGAFLVGPQFSNMHMKLVQGLIDIPGKIPIIS----- 226
DB 70 SQRLDAVARMPSLLELYLSYNCNFMGLIAGPLCSYDYITTFIEGRSYHTLGGSECKEETQ 129
QY 227 -----IIPAL-----KRLSLGLFVLVGYTLSPHITEDYLLTDEDYD 262
DB 130 YERTYAIKCKVMRFIMSLYSWYXAXVQKLLVGLSLFLHLICTT-LPVEYNIDEHQ 188
QY 263 NHPRF-FECMTMLMGKVLKYVTCVLTEGVCILTLGLGNGEPEKKAADACANMKV 321
DB 189 ATASMPRTIYLYLSLAAAPRYPAWTLADAINNAAGFRGYDENGAARMDLSNLR 248
QY 322 WLFETNPRFTGIASFENINTNAAWVARYIFKRLKF 355
DB 249 QQIEKSTSFKMFPLDNWNIQTALMPKRCYCIERTSF 282

RESULT 5
US-09-864-761-33415

Sequence 33415, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wenheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Aecmca-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

PRIOR FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Anncmax Sequence Listing Engine vers. 1.1

SEQ ID NO 33415

LENGTH: 58

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AC006512.12

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.5

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.6

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.8

OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.6

OTHER INFORMATION: EXPRESSED IN BETA LIVER, SIGNAL = 11

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.9

OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2

OTHER INFORMATION: EST HUMAN HIT: W26367.1, EVALUATE 3.00e-26

OTHER INFORMATION: SWISSPROT HIT: P43288, EVALUATE 4.50e+00

US-09-864-761-33415

Query Match 8.3%; Score 215; DB 9; Length 58;

Best Local Similarity 89.4%; Pred. No. 3.3e-13;

Matches 42; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 391 VIVERQARLIQESPTLSKAATVLOPFYLVQOTIHWLMFGYSMT 437

DB 12 LILPQARLIQESPTLSKLAITVLOPFYLVQOTIHWLMFGYSMT 58

RESULT 6
US-10-205-219-192

Sequence 192, Application US/10205219

Publication No. US20030138803A1

GENERAL INFORMATION:

APPLICANT: Warner-Lambert Company

APPLICANT: Lee, Kevin

APPLICANT: Dixon, Alistair

APPLICANT: Brooksbank, Robert

APPLICANT: Pinnock, Robert

TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain

FILE REFERENCE: W-1-A-018200

CURRENT APPLICATION NUMBER: US/10/205,219

PRIOR FILING DATE: 2002-07-24

PRIOR APPLICATION NUMBER: GB 0118354.0

PRIOR FILING DATE: 2001-07-27

NUMBER OF SEQ ID NOS: 197

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 192

LENGTH: 342

TYPE: PRT

ORGANISM: Rattus norvegicus

FEATURE:

OTHER INFORMATION: BBI

US-10-205-219-192

Query Match 5.8%; Score 150; DB 12; Length 342;

Best Local Similarity 24.5%; Pred. No. 5.9e-06;

Matches 67; Conservative 35; Mismatches 11; Indels 50; Gaps 11;

QY 226 SIIPALKRL-----SLGIFVLVGYTLSPHITEDYLLTD--YDNHPRFRCMYML-INGKF 279

DB 57 SLRPLLRAMPAPLPGILF-----LLSSHLPLEAIVEDAFYARPLPARLFYMPVFFAF 111

Qy 280 VLKRYTCVLTEGVCILATGLGFNGFEERKKA-----KWDACA 317
Db 112 RMRFFVY-AMIAECCGCIAGFCAYPAARAGAGGPTLCCPPSPSEKASLEVDYETIR 170
Qy 318 NKKWMLFFENPRFTGTIASFNINTNANVARYIFKRLKFKGNKLSGLSLPLALMHGH 377
Db 171 NIDCVSTDCVVRDGMRYMNMVOMWLAQYIKSNP-ARSYVLRATMTLMSAYWHGLH 229
Qy 378 SGVLVCFQMEFLIVIERQAAARLIQSPPLSKLAATTVLQPFYVLVQTIHMLF-----WG 433
Db 230 PGYVLSFLTIPCLIAAE---GRL--BSALGRLS-----PGQKAMDVNHFELKRAVD 278
Qy 434 YSMTAFCLFTMDKMLKVYSIYFLGHIFFLSLL 466
Db 279 YVCMGFVLLSLADTLRYMASIYFCIHFLALAL 311

RESULT 7
US-09-821-812-7
; Sequence 7, Application US/09821812
; Publication No. US20030166520A1
; GENERAL INFORMATION:
; APPLICANT: Lin. BiaoYang
; TITLE OF INVENTION: Androgen Regulated Prostate Specific
; FILE REFERENCE: P-IS 4373
; CURRENT APPLICATION NUMBER: US/09/821,812
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-821-812-7

Query Match 4.8%; Score 123.5; DB 12; Length 141;
Best Local Similarity 31.1%; Pred. No. 0.0063;
Matches 32; Conservative 19; Mismatches 45; Indels 7; Gaps 3;
Qy 118 FCFQMAVL---LAGYYRATGNVYDIKMTMPCVLTLLKLGAVDYFDG--GKQNSLSS 171
Db 11 FVFALGYLTVCCVTRVYIFDYQYSADFSGPMWITTKTSLACEIHDMFRKDB-ELTS 69
Qy 172 EQQXYAIRGVPSLLEVAGSFYFAGFLVGPQSPMNHMYKLVOG 214
Db 70 SGORDLAVRMPSLLEVLSTNCNFMGLAGPLCSYNDYITFIISG 112

RESULT 8
US-10-287-274-375
; Sequence 375, Application US/10287274
; Publication No. US20030181408A1
; GENERAL INFORMATION:
; APPLICANT: Foreysch, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERETO
; FILE REFERENCE: ELITRA 008DVI
; CURRENT APPLICATION NUMBER: US/10/287,274
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: US 09/711164
; PRIOR FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 375
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-287-274-375

Query Match 4.3%; Score 111; DB 12; Length 392;
Best Local Similarity 21.9%; Pred. No. 0.039;
Matches 93; Conservative 49; Mismatches 131; Indels 152; Gaps 22;
Qy 79 LSIATFNQNGQLYHLLCIVLQFLRLMGRTTAVLTFCQMAVLAGYYTA-TGNY 137
Db 72 LAMAVYVNRQGYHPL---VBPALASLFG-----YSLGGSITTDVGRY 113
Qy 138 DIKWTMPHCVLTLKLGAVDYFDGKQNSLSSBQO-----KVAIRGVPSLLEVAG 189
Db 114 ---WMLPYF-----YIPGHFNVSUVEFTAVCMTHIYIGVWMLLEFAPALFERLG 158
Qy 190 FSYFYGAFLVGPQSPMNHMYKLVOGELIDIPKIPNSIIPALKRSLGLFLVAGYTLSP 249
Db 159 WKV-----SLQRLNKVMFFIALGALLP 181
Qy 250 HTTE---DYLLTEDYDNHPEW-----FRCMYMLWGKFLVLYKVTCLVTEGVCIL 297
Db 182 TWHQSMGSLMISAGYKHPMQSYEMLPFLSLTAFTMGFSI-----VIFGSLVQ 233
Qy 298 TGLGFNGFEERKAKWACAMKMWLFETNPRFTGTIASFNINTNANVARYIFKRLKFLG 357
Db 234 AGLRNGFDEKS-----LF---VKLTITIS-----VLALFIVLR-G 267
Qy 358 NKELSQGLSLFLALMHGHSGLVLCFQMEFLIV---IVERQAAARLIQSPPLSKLAAT 414
Db 268 ELIYRDKLSLAPAG-----DFYSVMFIEVLMLFPLVLRVAKLRNDSRLF-LSALS 320
Qy 415 VLQPFYVIVQOTIHLFMNGYSMTAF-----CLF-TWPKMLKV-YKSIYFLGHIFFLSL 465
Db 321 AL-----LGCATWR-LTVSLVAFNPGGYAFYFPWELLISIGFVAIEICAYIVLRL 372
Qy 466 LFLIP 470
Db 373 LPILP 377

RESULT 9
US-10-202-824-6
; Sequence 6, Application US/10202824
; Publication No. US20030176648A1
; GENERAL INFORMATION:
; APPLICANT: Wood, John N.
; APPLICANT: Akopian, Armen N.
; TITLE OF INVENTION: Ion Channel
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESS: ZENECA Pharmaceuticals
; STREET: 1800 Concord Pike, P.O. Box 15437
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/202,824
; FILING DATE: 26-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,656
; FILING DATE: 24-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hohenschutz, Liza D.
; REGISTRATION NUMBER: 33,712
; REFERENCE/DOCKET NUMBER: PHW.70086
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 886-7466
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:

LENGTH: 2132 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-202-824-6

Query Match 4.1%; Score 105; DB 12; Length 2132;
Best Local Similarity 19.4%; Pred. No. 1.4;
Matches 113; Conservative 74; Mismatches 194; Indels 200; Gaps 28;

QY 7 GDEGVVALAGVLOGSGFQEE-----LSLNKATSLGASQEARL-----TISIF----- 49
DB 565 GEEGQGVPTSELTGTGAPGEPALDTTGQKSPLSAGYLNEPFRAGRAMSVSMTSVIEL 624
QY 50 ---LGYEPAL--FYRHYLFYK-----ETYLHLFH-----PTGSIAYFNGQLYSHL 94
DB 625 EESKLCPCPCLISFQKLLIMECCPKMKFKMALFELVTDPAELTIT----- 672
QY 95 LCIVLQPLILRL---MGRITVAULT-----FCFQMAVLLAG---YYTATGNYDI 139
DB 673 LCIVNTVFMMMEHYPMTDAPDMLQAGNIVFTVFETMEMAFKIAFDPIYY-----FQK 727
QY 140 KATMHCVL-FLKLGAVDFDGGKDNSSLSSEQKTAIRGVPSELLEVAGSFYFGAPL 198
DB 728 KMNIFDCVIVVLSLELS-----ASKKGSLS-----VLNSLRALDPTTGQKSPLSAGY 775
QY 199 VGPOPSNMHYMKLVQ-----GELIDIPGKIPNSII-----PALKRLSLGIFY 240
DB 776 LNEPRRARMASVSVSMTSVIELSESKLCPCCLISFQKLLIMECCPKMKFKMALFE 835
QY 241 LVGYLLSPHIT-----EDYLLTEDYDNHPFWRMVLIMGKEFLYKYVTCW 288
DB 836 LVTDPAELTITLCIVNTVFMMMEHYPMTDAPD-----ANLQAGNIVFTVFETME 886
QY 289 LVTEVCILTLGFGFGEKAKAKDA--CANMKTWLEFETNPRFGTITASF----- 337
DB 887 MAFK--IIADPYYYFQK---KMNIFDCVIVVLSLELSASKKGSLSVLSRLRLRVF 939
QY 338 ---NINTNMAVARYIFKRLKFLGNKELSGSLFLAL----- 372
DB 940 KLASWPLNT---LITIGNSVGLGNLFTLAIIVTIPALVQKQLSSEDCGRKQDVS 996
QY 373 -WHGLSGYLVG--FQMEFLIVVERQARLLOESPFLSKLAITVLQFPYLVQOTIH-- 428
DB 997 VMNGEKLAMHMDPFHSLV-----FRLLCGEMINM 1029
QY 429 WLFNGYSMTAFCLFTWDMKLVKYSIYFLGHIFFLSLFL 469
DB 1030 WVCMEVSQKSICLI-----LFLTVMLGNLVLNLFIAL 1063

RESULT 10
US-09-815-242-13958
Sequence 13958, Application US/09815242

GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 13958
LENGTH: 613
TYPE: PRT
ORGANISM: Salmonella typhi
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(613)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13958

Query Match 3.9%; Score 101; DB 9; Length 613;
Best Local Similarity 17.1%; Pred. No. 0.64;
Matches 110; Conservative 83; Mismatches 185; Indels 264; Gaps 29;

QY 22 GFQSLNKK-----LATSIGASQEARLITISFLGPFALFYHYLYKET----- 67
DB 14 GFVLAISRGWSNLSATIGVSGVGLAALVTAIVGDF--FANGKQAFSQPLTWMSVG 71
QY 68 -----YLHLPHTE-----TGLSIAYFNGQLYSHLSCI 97
DB 72 NFNIGFNLVDGLSLTMLSVTVGPFILHMFASVYMGEEYS--RFAYITLFLASNVVL 130
QY 98 VL--QFLIRLMGRTITAVLTTCFQMAVLLAGVYTT----- 132
DB 131 VLSNLLMLMYLMEGVGL-----CSYLLIGFYSDPKGAAMKAFVTVRGDVFIAF 183
QY 133 -----ATGNDIKRTMHCVLTLKLGAVDFDGGKDN 167
DB 184 ALFLYNELGLTNRENVETAPAFADGNMLMA-----TMLLGAIV-----GKSAQ 232
QY 168 SLSEQKQVAIRG---VPSILEVAGSFYFGAPLVGPOPSNMHYMKLVQSGELIDIPGKIP 224
DB 233 LPLOTWADAMAGTPPSALIHAAITM--VTAGVYLI---ARTHGLFLMTEPEILVLG-- 284
QY 225 NSIIPALKRLSLGSLFVL-----GYTLLSPHITTEDYLLTEDYD----- 262
DB 285 --IIGAITLVWAGPAAVQVDIKRVLAYSTMSQIGWFLA-----LGQAMDAALFHL 335
QY 263 -NPPWFRCMYMLIWGKFLV-----YK-----YVTQWLVTEGVCILTGFGFN 303
DB 336 MTHAF--FKALLFLASGVIILACHHEONIPIKQGLRKSIPLYVACFLVGAALSAIPVTA 394
QY 304 GFEEKGAKMDACANMKVWLEFETNPRFGTITASNINTNMAVARYIFKRLKFLNKELSQ 363
DB 395 GFESKDEILAGANANGHI-----NLMVAGLVGAR--MTSLYTFPMIP----- 434
QY 364 GLSLFLALMG---LHSGYLVCFQMEFLIVVERQARLLOESPFLSKLAITVLQOP 418
DB 435 -----IVFHGKXIHAAHGKXITHHLPLIVLMI-----LSGFVGAALVVP 474
QY 419 FYIVLQOTIH-----WLFNG-----YSMTAFCLFTW 444
DB 475 LQGVLPQTEILAHGRVLTLEITSGVAIAGIXIINAMLMGLRVLVTSIAXSAPGHLGWT 534
QY 445 DKMLKVKYSIYFLGHIFFLSLFLPIYIKHAMVPRKEXLKKM 486
DB 535 --WYNAM-----GFDWLYDKVYKPFVLGIMLKRDPNLAL 568

RESULT 11

```
US-09-769-787-119
; Sequence 119, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 119
; LENGTH: 263
; TYPE: PRF
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-119

Query Match      3.9%; Score 99.5; DB 11; Length 263;
Best Local Similarity 21.6%; Pred. No. 0.29;
Matches 59; Conservative 40; Mismatches 99; Indels 75; Gaps 11;

QY 53 PALAFRHHLYFKETYLHLHFTFGLSIAYNF-----GNQLYHSLLCIVLOF 101
DB 4 PLTFRRHRKRVVLLTLFLALVYRDVLTFFFDIADLAKKFDQAIKNDLKSALDF 63
QY 102 LILRLMGRITVAVLTFCFQMAVILAGYVYATAGNDIKMTMPCV-----TLTK-- 151
DB 64 RLQGF-----LGRYQSFIPIIIVLGLFGYIELKXVNLISGRVSYGKRLQVLA 119
QY 152 -----LIGLAVDYFDGKQD--NSLSSEQOKYALNGVPSL--EVAGFSYFYG 195
DB 120 SIPCLLYLVTLIAITTYFPGFSPDGNSLFSDS-----GLQRLDGEIKSYLFFTC 174
QY 196 AFLVGPQFSNMHYMKVQGELEIDIPKINSIIPAKRISLGL-----FYLGYT-- 245
DB 175 VLLIGIFNAIYFL-----QIVDYGVNVRSAITTYIMFLMSMLYSALPYMVPMTSL 229
QY 246 -----LSPHI--TEDYLLTEPDYDNH 264
DB 230 MQASVGDVSLMKLFTYILYIVYMWLEKXEDN 262

RESULT 12
US-09-781-880-5
; Sequence 5, Application US/09781880
; Publication No. US20030166042A1
; GENERAL INFORMATION:
; APPLICANT: Glucksman, Maria Alexandra
; APPLICANT: Siles-Santiago, Immaculada
; TITLE OF INVENTION: No. US20030166042A1e1 Seven-Transmembrane
; TITLE OF INVENTION: Proteins/G-Protein Coupled Receptors
; FILE REFERENCE: 35800/208932
; CURRENT APPLICATION NUMBER: US/09/781,880
; CURRENT FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/182,061
; PRIOR FILING DATE: 2000-02-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 663
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-781-880-5

Query Match      3.9%; Score 99.5; DB 12; Length 663;
Best Local Similarity 21.1%; Pred. No. 0.99;
Matches 108; Conservative 62; Mismatches 196; Indels 147; Gaps 23;
```

```
QY 24 OELSLNKLTATSGASEQALRLI-----ISIFGYPFALFYRHYLFYKETYLI--HLF 73
DB 116 EPLSNRRFTTALIGLVVCTICSCVMKTKQJMLBSAHMLPLARCLVPLETIYINKFA 175
QY 74 HTFTGLSIAYF-----NFGNQLYHSLLCIV---LOFLILRLMGRITVAVLTTCF 120
DB 176 MIFTLGVLVYFGLSNLVPVNLAKSAYRELQVVEVYGLALGSLMNQVLVPVLFVFWF 235
QY 121 QMAYLLAGYTTATGN-----YDKMTMPCVLT--LKLGLAVDYPPGQDQNSLS 170
DB 236 LVLFALQIYSYFSTRDQASRERLLFELTSLIAECCSPYSLGLVFTV----- 284
QY 171 SEQQYAIKGVPSLLEVAGFSYFGAFVAGFQFSNMHYMK-----LVQGELEIDIPGK 222
DB 285 -----SFVALGVUTLCKF--YIQGRAPFNDP--AMNRGTCGYTLILLAVGTGLIEL--- 333
QY 223 IPNSIIPALKRISLGLFYLGVYTLSP--HITEDVLLTEDVD--NHPFW--FR---CMYML 274
DB 334 ---QVHRAFLSLITLFIIVASIIQSMLEIADPIVLGASRDXSLWGHFRAVSLCLPFL 390
QY 275 IWGKVLKYVTCMLVTGVCILVGLGPNGEFEKQKAKMDACAMKXVLFETNRFCTI 334
DB 391 VFPAYMAYM-----ICOFFHMDP-----WLL-----II 413
QY 335 ASFNINTAMVARYIFKRLKFLGNKELSGSLFLALMGLHSGYLVCQFQMEFLIYVE 394
DB 414 ISSSLITSL-----QVLGTFLI-----YVFWMEERFKEVE 445
QY 395 ROARLIGESPPLSLAIAIVLQPPYLVQOTI--HMLFMGYSMTAFCLFTWDMKLYKK 452
DB 446 NMDDVIYVNGTYRLLEFLVALCVAVGVSESTIRQEWTVWG-SMIIFHSYNNVWLBRAQ- 503
QY 453 STYFGLHIFPLSLFILPYHKAWPRKEXKK 485
DB 504 ---LGMKSFLLRDAVNMKIKSLPLATKEOLEK 532

RESULT 13
US-09-815-242-5694
; Sequence 5694, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5694
```

```

; LENGTH: 587
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5694

```

Query Match	3.7%	Score 96.5;	DB 9;	Length 587;
Best Local Similarity	21.0%;	Pred. No. 1.6;		
Matches	78;	Conservative	52;	Mismatches 138;
			Indels	103;
			Gaps	17

QY	66	ETLLIHFFHTFETLSIAVFNFGOLVHSLCIVLOFLLRLRMGRITAVLTPFCFOAYL	125
Db	40	DTFFVVISGYLISLLI-----SEYRYKQIDULLEFWFKRK-LRIIPAVLFLICVLMTF	92
QY	126	L-----AGYYTATGNYDIKMTMPCHVLTLLKLGADYDPGDKONSLS	170
Db	93	LIFKEPELLIQMKRDAIAAIFYVSNWY-----ISQNDYFN-----	128
QY	171	SEQOKAINGVBEILLEVAGSFYFYGAFLVGPFSNMHMYKVGELDIDPKIKNSITPA	230
Db	129	---GPAIEPLKLMWLAIEEOFYLLFLPVITFELLHREF-----PRNIQT	170
QY	231	L---KELSLGFLVLYGVTLLSPHIED-----YLLTEDDNHPMPWRGMYMLIWGKFLYK	283
Db	171	LFLVSLISIGLMTVIHF-----ITGDSRYVFGT-DTRLQTLIGCLIAETWPPFALK	223
QY	284	YVTCMLVTBEGVCLITGLGNGFEEEKGAKAMDACAMMKWLEPETHPRF-----	337
Db	224	DISKTIIVS---LDIIGISGFVLTMLFFIVGDQDMY--NGGFIIISFATLFLIAI	276
QY	338	NITNNAWARYI-FKRLFLGNKEISQGLSLFLALMH-----GLHSGYLCFQNEFLIV	392
Db	277	AVHPSSIFAKFLSMKPELLIIIGRRYS-----LYLMHYPIIVFNVSYYVGQIPVYVI	329
QY	393	VERQARLIQOE	403
Db	330	ITLTLTALMAE	340

```

RESULT 14
US-09-815-242-12667
Sequence 12667, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyekind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815, 242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191, 078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206, 848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207, 727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242, 578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253, 625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257, 931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269, 308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: PASESEQ for Windows Version 4.0
SEQ ID NO 12667

```

```

-; LENGTH: 603
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12667

```

Query Match	3.7%;	Score 96.5;	DB 9;	Length 603;
Best Local Similarity	21.0%;	Pred. No. 1.7;		
Matches	78;	Conservative	52;	Mismatches 138;
				Indels 103;
				Gaps 17;

[illegible]

```

RESULT 15
US-09-731-872-287
/ Sequence 287, Application US/09731872
/ Patent No. US20020102604A1
GENERAL INFORMATION:
/ APPLICANT: Dumas Milne Edwards, Jean Baptiste
/ APPLICANT: Bougueleret, Lydie
/ TITLE OF INVENTION: FULL-LENGTH HUMAN CDNA5 ENCODING POTENTIALLY SECRETED PROTEINS
/ FILE REFERENCE: 78 US3 REG
/ CURRENT APPLICATION NUMBER: US/09/731,872
/ CURRENT FILING DATE: 2000-12-07
/ PRIOR APPLICATION NUMBER: US 60/169,629
/ PRIOR FILING DATE: 1999-12-08
/ PRIOR APPLICATION NUMBER: US 60/187,470
/ PRIOR FILING DATE: 2000-03-06
/ NUMBER OF SEQ ID NOS: 482
/ SOFTWARE: Patent.pm
/ SEQ ID NO 287
/ LENGTH: 286
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: SIGNAL
/ LOCATION: -48...-1
US-09-731-872-287

```

Query Match	3.7%	Score 95.5	DB 10	length 286
Best Local Similarity	22.1%	Pred. No. 0.78		
Matches	76	Conservative	39	Mismatches 102; Indels 127; Gaps 19
QY	161	DEGKDONSLSSEQOKAIRGVPVSLLEVAGSFIFYGAPLVGPFSNN---	HYMKLQYQELI	217
Db	6	DGGTSESIFPDL--YASMGIRSLTMVAGPVPYVILCVFVVCQQLSSLNATYKSLVAREV		62

```

Qy 218 --DIPKIPNSIIPALKRSLGLFVLVGYTLSPHITEDYLLTEDYDNHPFWRFCMYMLI 275
Db 63 FMDLAA---TRAVFGVQSTAAGLMALIG-----DVLHAD-----KARGQON 101
Qy 276 WGFVLYKYVTCWLVTEGVCIITGLGPNNGFEKKAQKADACAMKVMLEFETNRPFTGTA 335
Db 102 W-----CWF-----HITTATGFCEFE-----NVAVHLSNLIFFITFDLFL----- 135
Qy 336 SFNINTNMAWARYIFRKLKFLG--NKEISQGLSLFLAL-----MHGLHSGYLIV 382
Db 136 -----VIHHLFAFLGLGCLVNLQGHYLAHTTLLLEMSPTTCVSMMLLKAGW-- 184
Qy 383 CFQMEFLIVIVERQARLIQESPTLSKLAITVLQPFYLVQOTIHMLF-----MGYSMT 437
Db 185 -----SESL-----FWKLNQWLMITHMFHCRMVLYTHMW 212
Qy 438 AFCLFTWDKMLKVKYSIYFLGH--IFPLSL---LFILPY-IHK 474
Db 213 WVCFWHMDGLV---SSLYLPHLTLFLVGLALTLTLINPYWTHK 252

```

Search completed: November 21, 2003, 14:21:38
 Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 14:16:43 ; Search time 46 Seconds
(without alignments)
1680.431 Million cell updates/sec

Title: US-09-938-803-6
Perfect score: 487
Sequence: 1 MASSAEGDEFTVALAGVLQ.....ILPYIKAMPREKELKME 487

Scoring table:
Gapop 60.0, Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

A Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	487	100.0	487	21	AAV69987 Human receptor-ase
2	487	100.0	487	22	AAAB3169 Amino acid sequenc
3	487	100.0	487	22	AAAB36584 Human FLEXHR-6 pro
4	303	62.2	424	22	ABG06198 Novel human diagno
5	123	25.3	128	21	AAAG00235 Human secreted pro
6	42	8.6	58	22	ABG47472 Human liver peptid
7	42	8.6	58	22	ABB27468 Human peptide #119
8	42	8.6	58	22	ABB32619 Peptide #125 encod
9	42	8.6	58	22	ABB18117 Protein #116 encod

10	42	8.6	58	22	AAW53449
11	42	8.6	58	22	AAW65826
12	42	8.6	58	22	AAW13688
13	42	8.6	58	22	AAW06088
14	42	8.6	58	22	AAW01438
15	42	8.6	58	23	ABG35460
16	13	2.7	267	22	ABG06197
17	9	1.8	47	22	AAW06401
18	9	1.8	276	22	AAU5415
19	9	1.8	556	20	AAI13453
20	9	1.8	556	22	ABB60158
21	8	1.6	86	7	AAW60954
22	8	1.6	86	12	AAW11874
23	8	1.6	86	23	ABJ05390
24	8	1.6	86	23	ABJ05395
25	8	1.6	86	23	ABJ05397
26	8	1.6	86	23	ABJ05404
27	8	1.6	86	23	ABJ05408
28	8	1.6	86	23	ABJ05411
29	8	1.6	90	22	AAW08079
30	8	1.6	90	22	AAW08328
31	8	1.6	111	7	AAW60957
32	8	1.6	246	21	AAW22433
33	8	1.6	278	21	AAW22432
34	8	1.6	310	21	AAW22431
35	8	1.6	317	22	AAU19246
36	8	1.6	512	20	AAW03256
37	8	1.6	609	23	ABP73457
38	8	1.6	947	21	AAW25105
39	8	1.6	1252	22	AAW66357
40	8	1.6	1280	23	AAW18969
41	8	1.6	1280	23	AAW18979
42	7	1.4	7	22	AAW44238
43	7	1.4	20	22	AAW44242
44	7	1.4	20	22	AAW30406
45	7	1.4	21	22	AAW04308

ALIGNMENTS

RESULT 1
AAV69987 standard; Protein, 487 AA.

AAV69987;

31-MAY-2000 (first entry)

Human receptor-associated protein from Incyte clone 2906971.

Human receptor-associated protein, HRP; Incyte clone 2906971;
cytostatic; immunomodulatory; antiinflammatory; cardiant; antiautemic;
antiarteriosclerotic; hepatotropic; antiautemic antirheumatic;
antiautemic; osteopathic; antiallergic; antidiabetic; dermatological;
neuroprotective; diagnosis; treatment; prevention; reproductive disorder;
cardiovascular; cell proliferative; autoimmune; inflammatory; allergy;
gastrointestinal; atherosclerosis; cirrhosis; leukaemia; cancer; AIDS;
arthritis; anaemia; asthma; dermatitis; diabetes; osteoporosis;
multiple sclerosis; irritable bowel syndrome.

Homo sapiens.

Key	Location/Qualifiers
Modified-site	/note= "Potential phosphorylation site"
Modified-site	183
Modified-site	/note= "Potential phosphorylation site"
Modified-site	149
Modified-site	/note= "Potential phosphorylation site"
Modified-site	235
Modified-site	/note= "Potential phosphorylation site"

BN WO200008155-A2.
 XX 17-FEB-2000.
 XX 06-AUG-1999; 99WO-US17777.
 XX 07-AUG-1998; 98US-0160065.
 PR 01-SEP-1998; 98US-0098703.
 XX
 XX (INCY-) INCYTE PHARM INC.
 XX Hillman JL, Yue H, Lal P, Tang YT, Gorgone GA, Guegler KJ,
 PI Corley NC, Baughn MR;
 XX WPI; 2000-205710/18.
 DR N-PSDB; AAF50889.
 XX
 XX New human receptor-associated proteins (HRAP) useful for the diagnosis,
 PT treatment and prevention of cell proliferative, autoimmune,
 PT inflammatory, reproductive, cardiovascular, and gastrointestinal
 PT disorders
 PS
 XX Claim 1; Pages 73-74; 99pp; English.
 XX
 CC The Present sequence is a human receptor-associated protein
 CC (HRAP) from Incyte clone 2906971 obtained from THYMOTUS CDNA library.
 CC This sequence is expressed in nervous, gastrointestinal
 CC and reproductive tissues. HRAP has cytostatic, immunomodulatory,
 CC antiinflammatory, cardiac, antihypertensive, hepatotropic,
 CC antirheumatic, antipneumatic, osteopathic, antiallergic, antianemic,
 CC antiasthmatic, antidiabetic, dermatological and neuroprotective
 CC activities. The present sequence is useful in the diagnosis, treatment
 CC and prevention of disorders associated with HRAP expression, especially
 CC cell proliferative, autoimmune/inflammatory, reproductive,
 CC cardiovascular and gastrointestinal disorders (e.g. atherosclerosis,
 CC arthritis, leukemia, cancer, AIDS, arthritis, allergies, anaemia,
 CC asthma, dermatitis, diabetes, osteoporosis, multiple sclerosis and
 CC irritable bowel syndrome).
 XX
 XX Sequence 487 AA;
 SQ
 Query Match 100.0%; Score 487; DB 21; Length 487;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 YLVQOTIHWLPMGYSMTAFCLFTMDKMLKVYSIYFLGHIFLLSTFLPYTHKAVPRK 480
 DB 421 YLVQOTIHWLPMGYSMTAFCLFTMDKMLKVYSIYFLGHIFLLSTFLPYTHKAVPRK 480
 QY 481 EKLKKME 487
 DB 481 EKLKKME 487
 XX
 XX RESULT 2
 XX AAB31669
 ID AAB31669 strandard; Protein; 487 AA.
 XX
 AC AAB31669;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Amino acid sequence of a human protein having a hydrophobic domain.
 XX
 XX Human; hydrophobic protein; secretory protein; membrane protein; sepsis;
 KW tumour inhibition; immune deficiency; autoimmune disorder; anaemia; burn;
 KW infectious disease; cancer; ulcer; periodontal disease; coagulation;
 KW Parkinson's disease; fertility; immune response; thrombosis.
 XX
 XX Homo sapiens.
 OS
 XX WO200104297-A2.
 PN
 XX 18-JAN-2001.
 PD
 PF 16-JUN-2000; 2000MO-UP03942.
 XX
 XX 08-JUL-1999; 99UP-0194359.
 PR
 XX (SAGA) SAGAMI CHEM RES CENT.
 PA (PROT-) PROTEGENE INC.
 XX
 PI Kato S, Kimura T;
 XX
 DR WPI; 2001-103081/11.
 XX
 DR N-PSDB; AAF25159, AAF25169.
 XX
 PT Isolated human proteins and polynucleotides are used in research and
 PT have activities including cell proliferation/differentiation activity,
 PT immune stimulating activity and receptor/ligand activity -
 XX
 XX Claim 1; Page 90-93; 151pp; English.
 XX
 CC The present sequence represents a human protein with hydrophobic domains.
 CC The protein possesses a hydrophobic domain and so is a secretory protein
 CC or a membrane protein. The protein is used as an antigen to prepare
 CC antibodies. The polynucleotide sequence is useful as a source of probes
 CC for genetic diagnosis. It is also useful for producing the protein
 CC in large quantities and for gene therapy. The eukaryotic cells are used
 CC for detecting the receptors or ligands corresponding to the protein and
 CC for detecting small novel pharmaceuticals. The antibodies are also used
 CC for detection, quantification and purification of the proteins. Both the
 CC protein and polynucleotide may be used in research or as nutritional
 CC sources or supplements. The protein may have cytokine and cell
 CC proliferation/differentiation activity, immune stimulating and cell
 CC activity, hematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity and tumour inhibition activity. It may therefore be used to
 CC treat immune deficiencies resulting from autoimmune disorders or
 CC infectious diseases, cancer, sepsis, anaemias, burns and ulcers,
 CC periodontal disease, Parkinson's disease, induce fertility, improve
 CC immune response and enhance coagulation or inhibit thrombosis.
 XX
 XX Sequence 487 AA;
 SQ
 Query Match 100.0%; Score 487; DB 22; Length 487;
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	MASSAEGDEGTVALAGVLOSQFQELSLNKATLSIGASQOARLIIISIFLGYPALFPRH	60
DB	1	MASSAEGDEGTVALAGVLOSQFQELSLNKATLSIGASQOARLIIISIFLGYPALFPRH	60
QY	61	YLFYKETYLHLFHTFTGLSIAYFNFGNQLYHSLCTIVLOFLIRLMGRITTAVALTTFCF	120
DB	61	YLFYKETYLHLFHTFTGLSIAYFNFGNQLYHSLCTIVLOFLIRLMGRITTAVALTTFCF	120
QY	121	QMAVLLAGYTTATGNDYIKMTMPHCYVLTLLIGLAVDPFGGKQONSLSSEQKXAIRG	180
DB	121	QMAVLLAGYTTATGNDYIKMTMPHCYVLTLLIGLAVDPFGGKQONSLSSEQKXAIRG	180
QY	181	VPSLLEVAGFSYFYGAFLVGPQFSNMHMKLVQSELIDIPKIPINSIIPALKRSLGIFY	240
DB	181	VPSLLEVAGFSYFYGAFLVGPQFSNMHMKLVQSELIDIPKIPINSIIPALKRSLGIFY	240
QY	241	LVGYTLSPHITTEDYLLTEDYDNHPFWRQMYLIGKFEVLYKYVTCMLVTEGVCILTGL	300
DB	241	LVGYTLSPHITTEDYLLTEDYDNHPFWRQMYLIGKFEVLYKYVTCMLVTEGVCILTGL	300
QY	301	GFNGFEKGAQKMDACANMKVLPETNPRFTGTIASFNINNAVARYIFPKLFLGNKE	360
DB	301	GFNGFEKGAQKMDACANMKVLPETNPRFTGTIASFNINNAVARYIFPKLFLGNKE	360
QY	361	LSQGLSLFLALMHGHSGLVCFQMEFLIIVERQOARLIOESPTLSKLAITVLOPFY	420
DB	361	LSQGLSLFLALMHGHSGLVCFQMEFLIIVERQOARLIOESPTLSKLAITVLOPFY	420
QY	421	YLVQOTIHLFMGYSMTAFCLFTWDKMLKVYKSIYFLGHIIFPLSLFLPIYIHKAMVPRK	480
DB	421	YLVQOTIHLFMGYSMTAFCLFTWDKMLKVYKSIYFLGHIIFPLSLFLPIYIHKAMVPRK	480
QY	481	EKLKQME 487	
DB	481	EKLKQME 487	

RESULT 3

ID AAB36584 Standard; Protein: 487 AA.

AAB36584;

09-MAR-2001 (first entry)

Human FLEXHT-6 protein sequence SEQ ID NO:6.

Human, FLEXHT, full-length molecules expressed in human tissue;
diagnosis: gene expression; genetic linkage; genetic variability;
antianemic; anticonvulsant; antiarteriosclerotic; immunomodulatory;
cytostatic; antitubercular; antiinflammatory; hepatotropic; antidiabetic;
anti-gout; antihypertoid; neuroprotective; antiarthritic; osteopathic;
antiparasitic; antineumatic; antitumor; gene therapy; anaemia; gout;
epilepsy; arteriosclerosis; atherosclerosis; developmental disorder;
cancer; immunological disorder; asthma; bronchitis; cirrhosis;
Crohn's disease; diabetes mellitus; Grave's disease; multiple sclerosis;
osteoarthritis; pancreatitis; rheumatoid arthritis; psoriasis;
ulcerative colitis.

Homo sapiens.
WO200070047-A2.
23-NOV-2000.
12-MAY-2000; 2000WO-US13299.
14-MAY-1999; 99US-0311894.
14-MAY-1999; 99US-0311937.
14-MAY-1999; 99US-0311940.

PA	(INCYTE GENOMICS INC.																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
----	-----------------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

QY 481 EKLKME 487
 DB 481 EKLKME 487
 RESULT 4
 ID ABG06198 standard; Protein; 424 AA.
 AC ABG06198;
 DT 13-FEB-2002 (first entry)
 DE Novel human diagnostic protein #6189.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 PN WO200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 PA (HYSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS70385.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensic, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 PS Claim 20; SEQ ID No 36557; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II), or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensic, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 424 AA;
 Query Match 62.2%; Score 303; DB 22; Length 424;
 Best Local Similarity 100.0%; Pred. No. 5.1e-282;
 Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 64 YKETYLILHFRFTGSLAIVFNGQLYHSLLCIVLQGLILRLMGRITTAULTFFCFOMA 123
 |||||
 |||||

DB 1 YKETYLILHFRFTGSLAIVFNGQLYHSLLCIVLQGLILRLMGRITTAULTFFCFOMA 60
 QY 124 YLLAGYTYTATGNVDIKWTMPHCVLTLKLIGLAVYFDGKDQNSLSSEQKXAIRGVPS 183
 DB 61 YLLAGYTYTATGNVDIKWTMPHCVLTLKLIGLAVYFDGKDQNSLSSEQKXAIRGVPS 120
 QY 184 LLEVAGFSYFVGAFLVGPQFSNMHYMKLVQGLIDIPKIPISIIIPALKRISLGFYLVG 243
 DB 121 LLEVAGFSYFVGAFLVGPQFSNMHYMKLVQGLIDIPKIPISIIIPALKRISLGFYLVG 180
 QY 244 YTLSPHITEDYLTEDYDNHFPFRCMVMILMGKFLVYKYVTCWLVTGVCILTLGLGN 303
 DB 181 YTLSPHITEDYLTEDYDNHFPFRCMVMILMGKFLVYKYVTCWLVTGVCILTLGLGN 240
 QY 304 GFEKGGAKAKMDACANMKYWLFTNRPFTGTIASFNINTNWAVARYIPKRLKLNKELISQ 363
 DB 241 GFEKGGAKAKMDACANMKYWLFTNRPFTGTIASFNINTNWAVARYIPKRLKLNKELISQ 300
 QY 364 GLS 366
 DB 301 GLS 303
 RESULT 5
 ID AAG00235
 ID AAG00235 standard; Protein; 128 AA.
 AC AAG00235;
 XX
 DT 06-OCT-2000 (first entry)
 DE Human secreted protein, SEQ ID NO: 4316.
 DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 XX
 OS Homo sapiens.
 PN EP1033401-A2.
 PD 06-SEP-2000.
 PF 21-FEB-2000; 2000EP-0200610.
 PR 26-FEB-1999; 99US-0122487.
 PA (GEST) GENSET.
 PI Dumas Malne Edwards J, Duclert A, Giordano J;
 DR WPI; 2000-500381/45.
 DR N-PSDB; AAC00241.
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 PS Claim 13; SEQ ID 4316; 71pp + CD-ROM; English.
 XX The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.
 XX

SQ Sequence 128 AA;
Query Match 25.3%; Score 123; DB 21; Length 128;
Best Local Similarity 100.0%; Pred. No. 9.8e-110;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MASSAEGDEGVVALAGVLSGFGQELSLINKLATSIGASEQALRLIISIFLGGPALFYRH 60
DB 1 MASSAEGDEGVVALAGVLSGFGQELSLINKLATSIGASEQALRLIISIFLGGPALFYRH 60
OY 61 YLFYKETLILHFFHFTGLSTAYFNGQLYHSLLCIVLPFLILMGRITTAULTTCF 120
DB 61 YLFYKETLILHFFHFTGLSTAYFNGQLYHSLLCIVLPFLILMGRITTAULTTCF 120
OY 121 QMA 123
DB 121 QMA 123
Db 121 QMA 123
RESULT 6
ABG47472
ID ABG47472 standard; Peptide; 58 AA.
XX ABG47472;
AC
XX
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver peptide, SEQ ID NO 26120.
XX
KM Human, liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KM hypercholesterolaemia; coronary heart disease.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX
PF 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00664.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488898/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analysing gene expression in human adult liver -
XX
XX Claim 27; SEQ ID NO 26120; 658bp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (1) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 1109 defined nucleotide sequences given in the
CC specification (or complements/fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult
CC liver. (1) may be used for predicting, measuring and displaying gene
CC expression in samples derived from human adult liver. The genes
CC identified may be involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
CC is associated with coronary heart disease. ABG47348-ABG5930 represent
CC human liver single exon encoded peptides of the invention.
CC Note: The sequence information for this patent does not appear in the
CC printed specification but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 58 AA;
Query Match 8.6%; Score 42; DB 22; Length 58;
Best Local Similarity 100.0%; Pred. No. 2.7e-32;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 396 QARLIQSPFLSKAAITVLPFYLVQQTTHMLFMGYSMT 437
DB 17 QARLIQSPFLSKAAITVLPFYLVQQTTHMLFMGYSMT 58
RESULT 7
ABB27468
ID ABB27468 standard; Peptide; 58 AA.
XX
XX ABB27468;
AC
XX
XX
DT 01-FEB-2002 (first entry)
XX
DE Human peptide #119 encoded by breast cell single exon nucleic acid probe.
XX
KM Human, microarray; single exon probe; gene expression; breast;
KM disease; cancer.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PF 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes,
XX useful for measuring gene expression in sample derived from human
XX breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 27; SEQ ID NO 10436; 327bp + sequence listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 58 AA;
 Query Match 8.6%; Score 42; DB 22; Length 58;
 Best Local Similarity 100.0%; Pred. No. 2.7e-32;
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 396 QAARLIOESPTLSKLAITVLOPFYLVQQTTHLWFMGYSMT 437
 DB 17 QAARLIOESPTLSKLAITVLOPFYLVQQTTHLWFMGYSMT 58
 RESULT 8
 ABB32619
 ID ABB32619 standard; Peptide; 58 AA.
 AC ABB32619;
 XX
 DT 04-FEB-2002 (first entry)
 XX
 DE Peptide #125 encoded by human foetal liver single exon probe.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 XX
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX
 PS Claim 27; SEQ ID NO 25254; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 58 AA;
 Query Match 8.6%; Score 42; DB 22; Length 58;
 Best Local Similarity 100.0%; Pred. No. 2.7e-32;
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 396 QAARLIOESPTLSKLAITVLOPFYLVQQTTHLWFMGYSMT 437
 DB 17 QAARLIOESPTLSKLAITVLOPFYLVQQTTHLWFMGYSMT 58
 RESULT 9
 ABB18117

ID ABB18117 standard; Protein; 58 AA.
 XX
 AC ABB18117;
 XX
 DT 23-JAN-2002 (first entry)
 XX
 DE Protein #116 encoded by probe for measuring heart cell gene expression.
 XX
 KW Human; gene expression; heart; microarray; vascular system;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200157274-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00666.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 XX
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488899/53.
 XX
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts -
 XX
 PS Claim 15; SEQ ID NO 19887; 530pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.
 CC By measuring gene expression, the probes are useful for predicting,
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 58 AA;
 Query Match 8.6%; Score 42; DB 22; Length 58;
 Best Local Similarity 100.0%; Pred. No. 2.7e-32;
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 396 QAARLIOESPTLSKLAITVLOPFYLVQQTTHLWFMGYSMT 437
 DB 17 QAARLIOESPTLSKLAITVLOPFYLVQQTTHLWFMGYSMT 58
 RESULT 10
 AAM53449
 ID AAM53449 standard; Protein; 58 AA.
 AC AAM53449;
 XX
 DT 05-NOV-2001 (first entry)
 XX
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 25554.
 XX

KW Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer.
 OS Homo sapiens.
 FN WO200157275-A2.
 XX
 XX
 PD 09-AUG-2001.
 XX
 XX
 PF 30-JAN-2001; 2001WO-US00667.
 XX
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483446/52.
 XX
 XX
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 XX
 XX
 PS Example 4; SEQ ID NO: 25554; 650pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancer. The present sequence is a protein encoded by one of
 CC the probes of the invention.
 XX
 SQ Sequence 58 AA;
 QY Query Match 8.6%; Score 42; DB 22; Length 58;
 DB Best Local Similarity 100.0%; Pred. No. 2.7e-32;
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 396 QAAALIQSPITSLKLAITVLOPFYVVOQTTHWLFMGYSMT 437
 17 QAAALIQSPITSLKLAITVLOPFYVVOQTTHWLFMGYSMT 58
 RESULT 11
 ID AAM65826 standard; Protein; 58 AA.
 XX
 AC AAM65826;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed probe encoded protein SEQ ID NO: 26132.
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma.
 OS Homo sapiens.
 XX
 XX
 PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX
 XX
 PF 30-JAN-2001; 2001WO-US00668.
 XX
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488900/53.
 XX
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 XX
 PS Example 4; SEQ ID NO: 26132; 658bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention.
 XX
 SQ Sequence 58 AA;
 QY Query Match 8.6%; Score 42; DB 22; Length 58;
 DB Best Local Similarity 100.0%; Pred. No. 2.7e-32;
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 396 QAAALIQSPITSLKLAITVLOPFYVVOQTTHWLFMGYSMT 437
 17 QAAALIQSPITSLKLAITVLOPFYVVOQTTHWLFMGYSMT 58
 RESULT 12
 ID AAM13688 standard; Protein; 58 AA.
 XX
 AC AAM13688;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Peptide #122 encoded by probe for measuring cervical gene expression.
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 OS Homo sapiens.
 XX
 XX
 PN WO200157278-A2.
 XX
 PD 09-AUG-2001.
 XX
 XX
 PF 30-JAN-2001; 2001WO-US00670.
 XX
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488901/53.
 XX
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 XX

PS Claim 27; SEQ ID No 18514; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes
CC (SENP: see A1110068-A1128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human Hela cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 58 AA;

Query Match 8.6%; Score 42; DB 22; Length 58;
Best Local Similarity 100.0%; Pred. No. 2.7e-32;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 QAARLIQESPTLSKLAITVLOPFYLLVOQTTHMLFMGYSMT 437
Db 17 QAARLIQESPTLSKLAITVLOPFYLLVOQTTHMLFMGYSMT 58

RESULT 13

AA026088
ID AAM26088 standard; Protein; 58 AA.

AC AAM26088;
XX

DT 17-OCT-2001 (first entry)
XX

DE Peptide #125 encoded by probe for measuring placental gene expression.
XX

KW Probe; microarray: human; placenta; antenatal diagnosis;
XX

KM genetic disorder.
XX

OS Homo sapiens.
XX

PA WO200157272-A2.
XX

PI 09-AUG-2001.
XX

PT 30-JAN-2001; 2001WO-US00663.
XX

PS 04-FEB-2000; 2000US-0180312.
XX
26-MAY-2000; 2000US-0207456.
PR
30-JUN-2000; 2000US-0608408.
PR
03-AUG-2000; 2000US-0632366.
PR
21-SEP-2000; 2000US-0234687.
PR
27-SEP-2000; 2000US-0236359.
PR
04-OCT-2000; 2000GB-0024263.
XX

PA (MOLE-) MOLECULAR DYNAMICS INC.
XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX

DR WPI; 2001-488897/53.
XX

PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX

PS Claim 27; SEQ ID No 26357; 654bp; English.
XX

XX The present invention relates to single exon nucleic acid probes (SENP:
CC see A111315-A1157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX

SQ Sequence 58 AA;

Query Match 8.6%; Score 42; DB 22; Length 58;
Best Local Similarity 100.0%; Pred. No. 2.7e-32;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 QAARLIQESPTLSKLAITVLOPFYLLVOQTTHMLFMGYSMT 437
Db 17 QAARLIQESPTLSKLAITVLOPFYLLVOQTTHMLFMGYSMT 58

RESULT 14

AA001438
ID AAM01438 standard; Protein; 58 AA.

AC AAM01438;
XX

DT 09-OCT-2001 (first entry)
XX

DE Peptide #120 encoded by probe for measuring human breast gene expression.
XX

KW Probe; human; breast disease; breast cancer; development disorder;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX

OS Homo sapiens.
XX

PA WO200157270-A2.
XX

PI 09-AUG-2001.
XX

PT 29-JAN-2001; 2001WO-US00661.
XX

PS 04-FEB-2000; 2000US-0180312.
XX
26-MAY-2000; 2000US-0207456.
PR
30-JUN-2000; 2000US-0608408.
PR
03-AUG-2000; 2000US-0632366.
PR
21-SEP-2000; 2000US-0234687.
PR
27-SEP-2000; 2000US-0236359.
PR
04-OCT-2000; 2000GB-0024263.
XX

PA (MOLE-) MOLECULAR DYNAMICS INC.
XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX

DR WPI; 2001-476286/51.
XX

PT Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX

PS Claim 27; SEQ ID No 10178; 322bp; English.
XX

XX The present invention relates to novel single exon nucleic acid probes
CC (see A110010-A1110067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in
CC a human breast sample, where the probe hybridizes at high stringency to a
CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosing
CC diseases of the human breast, particularly those diseases with polygenic
CC aetiology. The diseases include: breast cancer, disorders of development,
CC inflammatory diseases of the breast, fibrocystic changes, proliferative
CC breast disease and non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 58 AA;

Query Match 8.6%; Score 42; DB 22; Length 58;
Best Local Similarity 100.0%; Pred. No. 2.7e-32;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 QAARLIQESPTLSKLAITVLOPFYLLVOQTTHMLFMGYSMT 437
Db 17 QAARLIQESPTLSKLAITVLOPFYLLVOQTTHMLFMGYSMT 58

RESULT 15
ABG35460
ID ABG35460 standard; Peptide; 58 AA.
XX
XX ABG35460;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human peptide encoded by genome-derived single exon probe SEQ ID 25125.
XX
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;
XX primary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX
XX Homo sapiens.
OS
XX WO200186003-A2.
PN
XX
XX 15-NOV-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00665.
PF
XX
XX 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
P1
XX WPI; 2002-114183/15.
DR
XX
XX Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX
XX Claim 27; SEQ ID No 25125; 634pp; English.
PS
XX
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung; comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the

CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
CC pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a peptide/protein
CC encoded by a single exon probe of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 58 AA;
Query Match 8.6%; Score 42; DB 23; Length 58;
Best local similarity 100.0%; Pred. No. 2.7e-32; Indels 0; Gaps 0;
Matches 42; Conservative -0; Mismatches 0;

Qy 396 QARLIQESPTLSKLAITVLPFYVVOQTTHWLFNGYSWT 437
DB 17 QARLIQESPTLSKLAITVLPFYVVOQTTHWLFNGYSWT 58

Search completed: November 21, 2003, 14:22:36
Job time : 47 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 14:20:54 / Search time 20 Seconds

(without alignments)
2341.707 Million cell updates/sec

Title: US-09-938-803-6

Sequence: 1 MASSAEGDEGTVALAGVLQ.....ILPYHKAVPRKELKKME 487

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: PIR 76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	9	1.8	556	2 A32466	numb protein - fru
2	9	1.8	1486	2 E96608	probable retroelem
3	8	1.6	86	1 NZPG	endozeptine - pig
4	8	1.6	86	2 S63593	acyl-coenzyme A-bi
5	8	1.6	86	2 S63594	acyl-coenzyme A-bi
6	8	1.6	86	2 S63595	acyl-coenzyme A-bi
7	8	1.6	87	1 NZBO	endozeptine - bovin
8	8	1.6	88	2 A57711	diazepam-binding i
9	8	1.6	380	2 C84710	hypothetical prote
10	8	1.6	504	2 S57589	probable membrane
11	8	1.6	508	1 A43713	calcium-dependent
12	8	1.6	547	2 D95337	probable manganese
13	7	1.4	54	2 H97854	hypothetical prote
14	7	1.4	55	2 D95140	hypothetical prote
15	7	1.4	86	2 S63592	acyl-coenzyme A-bi
16	7	1.4	87	1 NZRT	endozeptine [valida
17	7	1.4	87	1 NZRT	endozeptine - rat
18	7	1.4	87	2 A60059	endozeptine - mouse
19	7	1.4	100	2 B98008	hypothetical prote
20	7	1.4	101	4 S59321	hypothetical prote
21	7	1.4	107	2 S64345	hypothetical prote
22	7	1.4	126	2 S54159	probable sodium po
23	7	1.4	134	2 H84947	NADH2 dehydrogenas
24	7	1.4	150	2 G84173	archaeal histone A
25	7	1.4	162	2 E82875	hypothetical prote
26	7	1.4	176	2 AG0336	probable mannose-t
27	7	1.4	188	2 F81808	hypothetical prote
28	7	1.4	194	2 G83915	hypothetical prote
29	7	1.4	198	2 AF1569	hypothetical prote

30	7	1.4	228	2 F71831	hypothetical prote
31	7	1.4	242	1 S65576	phosphate uptake r
32	7	1.4	255	2 AE2740	outer membrane pro
33	7	1.4	256	2 T43869	phosphate transpor
34	7	1.4	259	2 C97521	27K outer membrane
35	7	1.4	260	1 OOCVCL	coat protein - tom
36	7	1.4	260	2 T26496	hypothetical prote
37	7	1.4	268	2 T27223	hypothetical prote
38	7	1.4	272	2 A72568	hypothetical prote
39	7	1.4	272	2 D64933	NAD synthase (EC 6
40	7	1.4	275	2 F90934	NAD synthetase [im
41	7	1.4	275	2 B85783	NAD synthetase, pr
42	7	1.4	275	2 AC0709	NH3-dependent NAD
43	7	1.4	276	2 G72292	glycerol-3-phospha
44	7	1.4	276	2 AH0460	conserved hypotet
45	7	1.4	292	2 T48327	hypothetical prote

ALIGNMENTS

RESULT 1
A32466
numb protein - fruit fly (*Drosophila melanogaster*)
C/Species: *Drosophila melanogaster*
C/Date: 21-Oct-1992 #sequence_revision 21-Oct-1992 #text_change 24-Sep-1998
C/Accession: A32466
R/Jemura, T.; Shepherd, S.; Ackerman, L.; Jan, L.Y.; Jan, Y.N.
Cell 58, 349-360, 1989
A/Title: numb, a gene required in determination of cell fate during sensory organ forma
A/Reference number: A32466; MUID:89324081; PMID:2752427
A/Accession: A32466
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-556 <UEM>
A/Cross-references: GB:M27815; NID:G158000; PID:G158001
C/Genetics:
A:Gene: FlyBase: numb
A:Cross-references: FlyBase: FBgn002973

Query Match 1.8%; Score 9; DB 2; length 556;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 360 ELSSQSL 368
DB 369 ELSSQSL 377

RESULT 2
E96608
probable retroelement polypeptide P25P12.89 [imported] - *Arabidopsis thaliana*
C/Species: *Arabidopsis thaliana* (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C/Accession: E96608
R/Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.H.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lurov, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, K.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A/Reference number: AB6141; MUID:21016719; PMID:11130712
A/Accession: E96608
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1486 <STO>
A/Cross-references: GB:AE005173; NID:G9954746; PIDN:AA09097.1; GSPDB:GN00141
C/Genetics:
A:Gene: P25P12.89

A:Map position: 1
C:Superfamily: retrovirus-related polypeptide

Query Match 1.6%; Score 9; DB 2; Length 1486;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 SIFLGYEFA 55
Db 752 SIFLGYEFA 760

RESULT 3
NZPG
endoneptine - pig
N:Alternate names: diazepam-binding inhibitor
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 26-Feb-1999
C:Accession: S00805; S36839
R:Chen, Z.W.; Agerberth, B.; Gell, K.; Andersson, M.; Mutt, V.; Oestenson, C.G.; Efendic
Burr, J. Biochem. 174, 229-245, 1998
A:Title: Isolation and characterization of porcine diazepam-binding inhibitor, a polypep-
tide release.
A:Accession: S00805; MUID:88254787; PMID:3289918
A:Reference number: S00805
A:Molecule type: protein
A:Residues: 1-86 <CHP>
R:Agerberth, B.; Boman, A.; Andersson, M.; Joernvall, H.; Mutt, V.; Boman, H.G.
Burr, J. Biochem. 216, 623-629, 1993
A:Title: Isolation of three antibacterial peptides from pig intestine: gastric inhibitor
A:Reference number: S36839; MUID:93387315; PMID:8375398
A:Accession: S36839
A:Molecule type: protein
A:Residues: 33-86 <ACE>
C:Comment: This protein may be the endogenous ligand of the benzodiazepine receptor.
C:Superfamily: endoneptine; endoneptine homology
C:Keywords: acyl-CoA binding; antibacterial; blocked amino end; intestine
F:1-86/Product: endoneptine #status experimental <MAT1>
F:18-74/Domain: endoneptine homology <NDZ>
F:33-86/Product: endoneptine (33-86) #status experimental <MAT2>
F:1/Modified site: blocked amino end (Ser) (probably acetylated) #status experimental

Query Match 1.6%; Score 8; DB 1; Length 86;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 308 KGRKAKMDA 315
Db 50 KGRKAKMDA 57

RESULT 4
S63593
acyl-coenzyme A-binding protein - turtle
C:Species: Testudines gen. sp. (turtle)
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 17-Mar-1999
C:Accession: S63593
R:Kragelund, B.B.; Hojrup, P.; Skou Jensen, M.; Karlskov Schjerling, C.; Juul, E.; Knud-
sen, J. Mol. Biol. 256, 187-200, 1996
A:Title: Fast and one-step folding of closely and distantly related homologous proteins
A:Reference number: S63592; MUID:96190960; PMID:8609609
A:Accession: S63593
A:Molecule type: protein
A:Residues: 1-86 <KRA>
C:Superfamily: endoneptine; endoneptine homology
C:Keywords: acyl-CoA binding
F:18-74/Domain: endoneptine homology <NDZ>

Query Match 1.6%; Score 8; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 308 KGRKAKMDA 315

Db 50 KGRKAKMDA 57

RESULT 5
S63594
acyl-coenzyme A-binding protein - mallard
C:Species: Anas platyrhynchos (mallard)
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C:Accession: S63594
R:Kragelund, B.B.; Hojrup, P.; Skou Jensen, M.; Karlskov Schjerling, C.; Juul, E.; Knud-
sen, J. Mol. Biol. 256, 187-200, 1996
A:Title: Fast and one-step folding of closely and distantly related homologous proteins
A:Reference number: S63592; MUID:96190960; PMID:8609609
A:Accession: S63594
A:Molecule type: protein
A:Residues: 1-86 <KRA>
C:Superfamily: endoneptine; endoneptine homology
C:Keywords: acyl-CoA binding
F:18-74/Domain: endoneptine homology <NDZ>

Query Match 1.6%; Score 8; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 308 KGRKAKMDA 315
Db 50 KGRKAKMDA 57

RESULT 6
S63595
acyl-coenzyme A-binding protein - chicken
C:Species: Gallus gallus (chicken)
C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C:Accession: S63595
R:Kragelund, B.B.; Hojrup, P.; Skou Jensen, M.; Karlskov Schjerling, C.; Juul, E.; Knud-
sen, J. Mol. Biol. 256, 187-200, 1996
A:Title: Fast and one-step folding of closely and distantly related homologous proteins
A:Reference number: S63592; MUID:96190960; PMID:8609609
A:Accession: S63595
A:Molecule type: protein
A:Residues: 1-86 <KRA>
C:Superfamily: endoneptine; endoneptine homology
C:Keywords: acyl-CoA binding
F:18-74/Domain: endoneptine homology <NDZ>

Query Match 1.6%; Score 8; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 308 KGRKAKMDA 315
Db 50 KGRKAKMDA 57

RESULT 7
NZBO
endoneptine - bovine
N:Alternate names: acyl-CoA-binding protein; diazepam-binding inhibitor
N:Containing: endoneptine, short form
C:Species: Bos primigenius taurus (cattle)
C>Date: 31-Dec-1988 #sequence_revision 30-Sep-1991 #text_change 22-Jun-1999
C:Accession: A26448; S23127; A25832; A27886; A32944
R:Webb, N.R.; Rose, T.M.; Malik, N.; Marguardt, H.; Shoyab, M.; Todaro, G.J.; Lee, D.C.
DNA 6, 71-79, 1987
A:Title: Bovine and human cDNA sequences encoding a putative benzodiazepine receptor 11
A:Reference number: A26448; MUID:87161236; PMID:2881742
A:Accession: A26448
A:Molecule type: mRNA
A:Residues: 1-87 <WMB>
A:Cross-references: EMBL:M15886; NID:g162978; PIDN:AAA30495.1; PID:g162979
R:Jensen, M.S.; Hojrup, P.; Rasmussen, J.T.; Knudsen, J.

Biochem. J. 284, 809-812, 1992
A:Title: Purification and characterization of variants of acyl-CoA-binding protein in th
A:Reference number: S23127; MUID:92322005; PMID:1622397
A:Accession: S23127
A:Molecule type: protein
A:Residues: 2-87 <JEN>
A:Experimental source: liver
R:Margardt, H.; Todaro, G.J.; Shoyab, M.
J. Biol. Chem. 261, 9727-9731, 1986
A:Title: Complete amino acid sequences of bovine and human endozepines: homology with re
A:Reference number: A25832; MUID:86278003; PMID:3525533
A:Accession: A25832
A:Molecule type: protein
A:Residues: 2-87 <MAR>
A:Experimental source: brain
R:Mikeelsen, J.; Hojrup, P.; Nielsen, P.F.; Roepstorff, P.; Knudsen, J.
Biochem. J. 245, 857-861, 1987
A:Title: Amino acid sequence of acyl-CoA-binding protein from cow liver.
A:Reference number: A27886; MUID:88024142; PMID:3663196
A:Accession: A27886
A:Molecule type: protein
A:Residues: 2-87 <MIK>
A:Experimental source: liver
R:Besman, M.J.; Yanagibashi, K.; Lee, T.D.; Kawamura, M.; Hall, P.F.; Shively, J.E.
Proc. Natl. Acad. Sci. U.S.A. 86, 4897-4901, 1989
A:Title: Identification of des-(Gly-Ile)-endozepine as an effector of corticotropin-depe
receptor.
A:Reference number: A32944; MUID:89296911; PMID:2544879
A:Accession: A32944
A:Molecule type: protein
A:Residues: 2-85 <BES>
A:Experimental source: adrenal cortex
C:Comment: This protein may take part in the modulation of GABA-ergic transmission in th
C:Superfamily: endozepine; endozepine homology
C:Keywords: acetylated amino end; acyl-CoA binding
F:2-87/Product: endozepine #status experimental <MAT>
F:2-85/Product: endozepine, short form #status experimental <MA2>
F:19-75/Domain: endozepine homology <ND2>
F:2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental

Query Match 1.6%; Score 8; DB 1; Length 87;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 KGAKWDA 315
DB 51 KGAKWDA 58

RESULT 8
A57711.
diazepam-binding inhibitor - laughing frog
C:Species: Rana ridibunda (laughing frog)
C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 13-Aug-1999
C:Accession: A57711
R:Uhrmann, I.; Plagevent, J.C.; Tostivint, H.; Rajmakers, R.; Tonon, M.C.; Conlon, J.
Proc. Natl. Acad. Sci. U.S.A. 91, 6899-6903, 1994
A:Title: Frog diazepam-binding inhibitor: peptide sequence, cDNA cloning, and expression
A:Reference number: A57711; MUID:94316605; PMID:8041717
A:Accession: A57711
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-88 <LIH>
A:Cross-references: GB:U09205; NID:g484097; PIDN:AAB60606.1; PID:g514280
C:Superfamily: endozepine; endozepine homology
F:20-76/Domain: endozepine homology <ND2>

Query Match 1.6%; Score 8; DB 2; Length 88;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 KGAKWDA 315
DB 51 KGAKWDA 58

DB 52 KGAKWDA 59

RESULT 9
C84710
hypothetical protein At2g30590 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84710
R:Lin, X.; Kaul, S.; Roundley, S.D.; Shee, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayem, L.; Tallon, L.;
Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617979
A:Accession: C84710
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-380 <STO>
A:Cross-references: GB:AE002093; NID:g1946360; PIDN:AAB63078.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g30590
A:Map position: 2

Query Match 1.6%; Score 8; DB 2; Length 380;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 SGFOELSL 28
DB 103 SGFOELSL 110

RESULT 10
S57589
probable membrane protein YMR221c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR959.03c
C:Species: Saccharomyces cerevisiae
C:Date: 19-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 19-Apr-2002
C:Accession: S57589
R:Skellton, J.; Churcher, C.M.
Submitted to the EMBL Data Library, June 1995
A:Reference number: S57587
A:Accession: S57589
A:Molecule type: DNA
A:Residues: 1-504 <SKE>
A:Cross-references: EMBL:Z49393; NID:g887599; PID:g887602; GSPDB:GN00013; MIPS:YMR221c
A:Experimental source: strain AB972
C:Genetics:
A:Gene: MIPS:YMR221c
A:Cross-references: SGD:S0004834
A:Map position: 13R
C:Superfamily: Saccharomyces cerevisiae probable membrane protein YMR221c
C:Keywords: transmembrane protein
F:12-28/Domain: transmembrane #status predicted <TM1>
F:65-81/Domain: transmembrane #status predicted <TM2>
F:93-109/Domain: transmembrane #status predicted <TM3>
F:127-143/Domain: transmembrane #status predicted <TM4>
F:190-206/Domain: transmembrane #status predicted <TM5>
F:306-322/Domain: transmembrane #status predicted <TM6>
F:349-365/Domain: transmembrane #status predicted <TM7>
F:374-390/Domain: transmembrane #status predicted <TM8>
F:399-415/Domain: transmembrane #status predicted <TM9>
F:467-483/Domain: transmembrane #status predicted <TM10>

Query Match 1.6%; Score 8; DB 2; Length 504;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 240 YLVGYTLL 247
DB 121 YLVGYTLL 128

RESULT 11

A:3713
C:Species: calcium-dependent protein kinase (EC 2.7.1.-) - soybean
C:Date: 03-Mar-1993 #sequence_revision 14-Jul-1994 #text_change 11-Jun-1999
C:Accession: A43713
R:Harper, J.F.; Sussman, M.R.; Schaller, G.E.; Putnam-Evans, C.; Charbonneau, H.; Harmon
Science 252, 951-954, 1991
A:Title: A calcium-dependent protein kinase with a regulatory domain similar to calmodulin
A:Reference number: A43713; MUID:91240279; PMID:1852075
A:Accession: A43713
A:Molecule type: mRNA
A:Residues: 1-508 <HAR>
A:Cross-references: EMBL:M64987; NID:g169930; PIDN:AAB00806.1; PID:g169931
C:Superfamily: calcium-dependent protein kinase; calmodulin repeat homology; protein kin
C:Keywords: ATP; calcium binding; EF hand; phosphotransferase; serine/threonine-specific
F:32-292/Domain: protein kinase homology <KIN>
F:40-48/Region: protein kinase ATP-binding motif
F:335-367/Domain: calmodulin repeat homology <EF1>
F:371-403/Domain: calmodulin repeat homology <EF2>
F:407-439/Domain: calmodulin repeat homology <EF3>
F:441-473/Domain: calmodulin repeat homology <EF4>
F:63/Active site: Lys #status predicted

Query Match 1.6%; Score 8; DB 1; Length 508;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 394 EROARLI 401
|||||
Db 132 EROARLI 139

RESULT 12

D95337
C:Species: Sinothizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: D95337
R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bowe
; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yen, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9683-9688, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinothizobium melilot
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: D95337
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-547 <KUR>
A:Cross-references: GB:AE006469; PIDN:AAK65262.1; PID:g14523713; GSPDB:GN00165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Phliher, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The complete genome of the legume symbiont Sinothizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Sma115
A:Genome: plasmid

Query Match 1.6%; Score 8; DB 2; Length 547;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LIGLAVDY 159
|||||
Db 472 LIGLAVDY 479

RESULT 13

H97854
C:Species: hypothetical protein RC1240 [imported] - Rickettsia conorii (strain Malish 7)
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: H97854
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
Science 293, 2093-2098, 2001
A:Title: Mechanisms of evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: H97854
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-54 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL03778.1; PID:g15620374; GSPDB:GN00173
C:Genetics:
A:Gene: RC1240

Query Match 1.4%; Score 7; DB 2; Length 54;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 AEGDEGT 11
|||||
Db 25 AEGDEGT 31

RESULT 14

D95140
C:Species: hypothetical protein SP1210 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: D95140
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: D95140
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-55 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75317.1; PID:g14972691; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1210

Query Match 1.4%; Score 7; DB 2; Length 55;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 LSGGLSL 367
|||||
Db 4 LSGGLSL 10

RESULT 15

S63592
C:Species: acyl-coenzyme A-binding protein - dog
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
C:Accession: S63592
R:Kragelund, B.B.; Hojrup, P.; Skou Jensen, M.; Karlskov Schierling, C.; Juul, E.; Knuds
J. Mol. Biol. 256, 187-200, 1996
A:Title: Fast and one-step folding of closely and distantly related homologous proteins
A:Reference number: S63592; MUID:96190960; PMID:8609609
A:Accession: S63592
A:Molecule type: protein
A:Residues: 1-86 <KGA>
C:Superfamily: endozepine; endozepine homology
C:Keywords: acyl-CoA binding

F:18-74/Domain: endozepine homology <NDZ>

Query Match 1.4%; Score 7; DB 2; Length 86;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 309 GKAKMDA 315

|||||

Db 51 GKAKMDA 57

Search completed: November 21, 2003, 14:24:33
Job time : 22 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: November 21, 2003, 14:17:14 ; Search time 17 Seconds

(without alignments)
1347.177 Million cell updates/sec

Title: US-09-938-803-6

Perfect score: 487
Sequence: 1 MASSAEGDEGTVALAGVLQ.....ILPYIKAMVPRKELKME 487

Scoring table: OLIGO

Gaped 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.8	556	1	NUMB_DROME
2	8	1.6	86	1	ACBP_BOVIN
3	8	1.6	86	1	ACBP_CHICK
4	8	1.6	86	1	ACBP_CHICK
5	8	1.6	86	1	ACBP_PIG
6	8	1.6	87	1	ACBP_RANR
7	8	1.6	283	1	HIS1_SULTO
8	8	1.6	310	1	SYB1_ARATH
9	8	1.6	380	1	WR21_ARATH
10	8	1.6	504	1	YME8_YEAST
11	8	1.6	508	1	CDPK_SOYBN
12	7	1.4	64	1	AMP2_MELGA
13	7	1.4	64	1	GLI3_CHICK
14	7	1.4	86	1	ACBP_CANFA
15	7	1.4	86	1	ACBP_HUMAN
16	7	1.4	86	1	ACBP_MOUSE
17	7	1.4	86	1	ACBP_RAT
18	7	1.4	107	1	YGLY_YEAST
19	7	1.4	126	1	CLIX_HUMAN
20	7	1.4	134	1	NUOA_BUCAL
21	7	1.4	228	1	YD31_HBLPJ
22	7	1.4	222	1	PHOU_PSEAE
23	7	1.4	256	1	PHOU_PSEAE
24	7	1.4	260	1	COAT_TYLCV
25	7	1.4	275	1	NADE_ECO57
26	7	1.4	275	1	NADE_ECOLI
27	7	1.4	275	1	NADE_SALTI
28	7	1.4	275	1	NADE_SALTI
29	7	1.4	294	1	CHTB_TOBAC
30	7	1.4	299	1	HEM3_RICPR
31	7	1.4	322	1	Y378_METUA
32	7	1.4	323	1	CITG_HAEDU
33	7	1.4	328	1	ZIPA_YERPE

34	7	1.4	333	1	MOSA_RHINE	Q07607 rhizobium m
35	7	1.4	353	1	FRSW_BUCAP	Q85913 buchnera ap
36	7	1.4	353	1	SOHB_HAEIN	P45315 haemophilus
37	7	1.4	354	1	CAHC_MOUSE	Q84185 mus musculus
38	7	1.4	355	1	CAHC_RABIT	Q96230 oryctolagus
39	7	1.4	369	1	DCHS_RHILLO	Q98407 rhizobium l
40	7	1.4	376	1	HIS8_SULSO	Q33770 sulfolobus
41	7	1.4	397	1	CD61_SULSO	Q98004 sulfolobus
42	7	1.4	419	1	RRAA_SALTI	Q842V3 salmoneilla
43	7	1.4	419	1	RRAA_SALTY	P27031 salmoneilla
44	7	1.4	443	1	RRG2_HUMAN	P22932 homo sapien
45	7	1.4	447	1	RRG2_MOUSE	P20787 mus musculus

ALIGNMENTS

RESULT 1

ID	NUMB_DROME	STANDARD	PRT	556 AA
AC	P16554			
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-AUG-1990 (Rel. 15, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	NUMB protein.			
GN	NUMB.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=89324061; PubMed=2752427;			
RA	Uemura T., Shepherd S., Ackerman L., Jan L.Y., Jan Y.N.;			
RT	"Numb, a gene required in determination of cell fate during sensory			
RT	organ formation in Drosophila embryos.";			
RL	Cell 58:349-360(1989).			
RN	[2]			
RP	STRUCTURE BY NMR OF 64-210.			
RX	MEDLINE=99061335; PubMed=9846878;			
RA	Li S.-C., Zwaalen C., Vincent S.J., McGlade C.J., Kay L.E., Pawson T.,			
RT	Forman-Kay J.D.;			
RT	"Structure of a Numb PTB domain-peptide complex suggests a basis for			
RT	diverse binding specificity.";			
RL	Nat. Struct. Biol. 5:1075-1083(1998).			
CC	-I- FUNCTION: NUMB IS REQUIRED IN DETERMINATION OF CELL FATE DURING			
CC	SENSORY ORGAN FORMATION IN DROSOPHILA EMBRYOS. IT FUNCTIONS IN			
CC	-I- SUBCELLULAR LOCATION: Nuclear.			
CC	-I- ALTERNATIVE PRODUCTS:			
CC	Event=Alternative initiation;			
CC	Comment=2 isoforms, Long/Zygotic (shown here) and			
CC	Short/Maternal, may be produced by alternative initiation;			
CC	-I- SIMILARITY: Contains 1 PID domain.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL; M27815; AAA28730.1; .			
DR	PIR; A32466; A32466.			
DR	PDB; 2NMB; 04-NOV-98.			
DR	PDB; 1DDM; 12-APR-00.			
DR	flyBase; FBgn0002973; numb.			
DR	GO; GO:0045180; C:basal cortex; NAS.			
DR	InterPro; IPR006020; PTB_PID.			
DR	Pfam; PF00640; PID; 1.			
DR	SMART; SM00462; PTB; 1.			

DR PROSITE; PS01179; PID: 1.
 KW Developmental protein; Nuclear protein; ATP-binding;
 Alternative initiation; 3D-structure.
 FT CHAIN 1 556 NUMB PROTEIN, ISOFORM LONG.
 FT INIT MET 42 556 NUMB PROTEIN, ISOFORM SHORT.
 FT NP BIND 42 42 FOR ISOFORM SHORT.
 FT DOMAIN 22 29 ATP (POTENTIAL).
 FT DOMAIN 25 57 ARG/LYS-RICH (BASIC).
 FT DOMAIN 81 208 PID.
 FT TURN 66 68
 FT HELIX 69 78
 FT STRAND 83 92
 FT HELIX 100 111
 FT TURN 112 113
 FT STRAND 119 122
 FT STRAND 129 132
 FT STRAND 138 138
 FT TURN 144 146
 FT STRAND 163 165
 FT TURN 167 168
 FT STRAND 172 179
 FT HELIX 188 194
 FT TURN 195 198
 FT TURN 200 201
 SQ SEQUENCE 556 AA; 60628 MW; 4FECAB9C98FEE71 CRC64;
 Query Match 1.8%; Score 9; DB 1; Length 556;
 Best Local Similarity 100.0%; Pred. No. 0.86;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 360 ELSQGLSL 368
 DB 369 ELSQGLSL 377

RESULT 2
 ACPB BOVIN
 ID ACPB BOVIN STANDARD; PRT; 86 AA.
 AC P07107;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Acyl-CoA-binding protein (ACBP) (Diazepam binding inhibitor) (DBI)
 DE (Endozepine) (EP).
 GN DBI.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87161236; PubMed=2881742;
 RA Webb N.R., Rose T.W., Malik N., Margardt H., Shoyab M., Todaro G.J.,
 RA Lee D.C.;
 RA "Bovine and human cDNA sequences encoding a putative benzodiazepine
 RT receptor ligand."
 RT DNA 6:71-79(1987).
 RL [2]
 RN [2]
 RP SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=86278003; PubMed=3525533;
 RA Margardt H., Todaro G.J., Shoyab M.;
 RT "Complete amino acid sequences of bovine and human endozepines.
 RT Homology with rat diazepam binding inhibitor."
 RL J. Biol. Chem. 261:9727-9731(1986).
 RN [3]
 RP SEQUENCE.
 RC TISSUE=Liver;
 RX MEDLINE=88024142; PubMed=3663196;
 RA Mikkelsen U., Hoejrup P., Nielsen P.F., Roepstorff P., Knudsen J.;
 RA "Amino acid sequence of acyl-CoA-binding protein from cow liver."
 RL Biochem. J. 245:857-861(1987).

RN [4]
 RP SEQUENCE.
 RC TISSUE=Liver;
 RX MEDLINE=92322005; PubMed=1622397;
 RA Jensen M.S., Hoejrup P., Rasmussen J.T., Knudsen J.;
 RT "Purification and characterization of variants of acyl-CoA-binding
 RL protein from bovine liver."
 RL Biochem. J. 284:809-812(1992).
 RN [5]
 RP STRUCTURE BY NMR.
 RX MEDLINE=92031512; PubMed=1931985;
 RA Andersen K.V., Ludvigsen S., Wandrup S., Poulsen F.M.;
 RT "The secondary structure in solution of acyl-coenzyme A binding
 RT protein from bovine liver using 1H nuclear magnetic resonance
 RL spectroscopy."
 RL Biochemistry 30:10654-10663(1991).
 RN [6]
 RP STRUCTURE BY NMR.
 RX MEDLINE=92389322; PubMed=1518047;
 RA Andersen K.V., Poulsen F.M.;
 RT "Three-dimensional structure in solution of acyl-coenzyme A binding
 RL protein from bovine liver."
 RL J. Mol. Biol. 226:1131-1141(1992).
 RN [7]
 RP STRUCTURE BY NMR.
 RX MEDLINE=93364192; PubMed=8358232;
 RA Andersen K.V., Poulsen F.M.;
 RT "The three-dimensional structure of acyl-coenzyme A binding protein
 RT from bovine liver: structural refinement using heteronuclear
 RT multidimensional NMR spectroscopy."
 RL J. Biomol. NMR 3:271-284(1993).
 CC - FUNCTION: BINDS MEDIUM- AND LONG-CHAIN ACYL-COA ESTERS WITH VERY
 CC HIGH AFFINITY AND MAY FUNCTION AS AN INTRACELLULAR CARRIER OF
 CC ACYL-COA ESTERS. IT IS ALSO ABLE TO DISPLACE DIAZEPAM FROM THE
 CC BENZODIAZEPINE (BDZ) RECOGNITION SITE LOCATED ON THE GABA TYPE A
 CC RECEPTOR. IT IS THEREFORE POSSIBLE THAT THIS PROTEIN ALSO ACTS
 CC AS A NEUROPEPTIDE TO MODULATE THE ACTION OF THE GABA RECEPTOR.
 CC - SUBUNIT: Monomer.
 CC - SIMILARITY: Belongs to the ACPB family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M15886; AAA30495.1; -
 CC PIR; A26448; NZBO.
 CC PDB; 2ABD; 15-JUL-93.
 CC PDB; 1ACA; 30-APR-94.
 CC PDB; 1HB8; 11-MAR-02.
 CC InterPro; IPR000582; Ac_coA_bind_prot.
 CC Pfam; PF00887; ACPB; 1.
 CC PRINTS; PR00689; ACOBABINDING.
 CC Prodom; PD351532; Ac_coA_bind_prot; 1.
 DR PROSITE; PS00880; ACPB; 1.
 KW Transport; Lipid-binding; Acetylation; 3D-structure.
 FT INIT MET 0 0
 FT MOD_RES 1 1 ACETYLTATION.
 FT HELIX 4 14
 FT HELIX 21 35
 FT TURN 36 36
 FT TURN 48 61
 FT HELIX 62 63
 FT TURN 66 83
 FT HELIX 84 86
 FT TURN 84 86
 SQ SEQUENCE 86 AA; 9913 MW; 6B46B7C6C8CDE17 CRC64;
 Query Match 1.6%; Score 8; DB 1; Length 86;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 KGKAKWDA 315
 |||||
 50 KGKAKWDA 57

Db

RESULT 3

ACBP_CHAVI STANDARD; PRT; 86 AA.

AC P82934;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Acyl-CoA-binding protein (ACBP) (EP).
 GN DBI

OS Chaelephractus villosus (South American armadillo).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Edentata; Dasypodidae; Chaelephractus.
 OX NCBI_TaxID=23080;
 RN [1]
 RP SEQUENCE, AND ACETYLATION.
 RC TISSUE=Harderian gland;
 RX MEDLINE=21240303; PubMed=11342056;
 RA Cavagnari B.M., Sterin-Speziale N., Affanni J.M., Knudsen J.,
 Santome J.A.;
 RT "Acyl-CoA-binding protein in the armadillo Harderian gland: its
 RT primary structure and possible role in lipid secretion."
 RL Biochim. Biophys. Acta 1545:314-325(2001).
 CC -1- FUNCTION: BINDS MEDIUM- AND LONG-CHAIN ACYL-COA ESTERS WITH VERY
 CC HIGH AFFINITY AND MAY FUNCTION AS AN INTRACELLULAR CARRIER OF
 CC ACYL-COA ESTERS.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SIMILARITY: Belongs to the ACBP family.
 DR HSSP: P07107; IACA.
 DR InterPro: IPR000582; Ac_coa_bind_prot.
 DR Pfam: PF00887; ACBP; 1.
 DR PRINTS: PR00689; ACOABINDINGP.
 DR PRODOM: PD351532; Ac_coa_bind_prot; 1.
 DR PROSITE: PS00880; ACBP; 1.
 KM Transport; Lipid-binding; Acetylation.
 FT MOD_RES 1 ACETYLATION.
 SQ SEQUENCE 86 AA; 9891 MW; 6D6803B10A2C8CD CRC64;

QY 308 KGKAKWDA 315
 |||||
 50 KGKAKWDA 57

Db

RESULT 4

ACBP_CHICK STANDARD; PRT; 86 AA.

AC O9PRL8;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Acyl-CoA-binding protein (ACBP).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Avea; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=9619060; PubMed=8609609;
 RA Kirgelund B.B., Hoejrup P., Jensen M.S., Schjerling C.K., Juul E.,
 Knudsen J., Poulsen F.M.;
 RT "Fast and one-step folding of closely and distantly related homologous
 RT proteins of a four-helix bundle family."
 RL J. Mol. Biol. 256:187-200(1996).

CC -1- FUNCTION: BINDS MEDIUM- AND LONG-CHAIN ACYL-COA ESTERS WITH VERY
 CC HIGH AFFINITY AND MAY FUNCTION AS AN INTRACELLULAR CARRIER OF
 CC ACYL-COA ESTERS (BY SIMILARITY).
 CC -1- SIMILARITY: Belongs to the ACBP family.
 DR PIR: S63595; S63595.
 DR HSSP: P07107; IACA.
 DR InterPro: IPR000582; Ac_coa_bind_prot.
 DR Pfam: PF00887; ACBP; 1.
 DR PRINTS: PR00689; ACOABINDINGP.
 DR PRODOM: PD351532; Ac_coa_bind_prot; 1.
 DR PROSITE: PS00880; ACBP; 1.
 KM Transport; Lipid-binding.
 SQ SEQUENCE 86 AA; 9645 MW; 5AC1824914DCEB3 CRC64;

QY 308 KGKAKWDA 315
 |||||
 50 KGKAKWDA 57

Db

RESULT 5

ACBP_PIG STANDARD; PRT; 86 AA.

AC P12026;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Acyl-CoA-binding protein (ACBP) (Diazepam binding inhibitor) (DBI)
 DE (Endocephaline) (EP) (Contains: DBI [32-86]).
 GN DBI.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Intestine;
 RX MEDLINE=88254787; PubMed=3289918;
 RA Chen Z.W., Agerberth B., Gell K., Andersson M., Mutt V.,
 Oerter J., Eftendic S., Barros-Soderling J., Persson B.,
 Joernvall H.;
 RT "Isolation and characterization of porcine diazepam-binding
 RT inhibitor, a polypeptide not only of cerebral occurrence but also
 RT common in intestinal tissues and with effects on regulation of
 RT insulin release."
 RL Eur. J. Biochem. 174:239-245(1988).
 RN [2]
 RP SEQUENCE OF 32-86.
 RC TISSUE=Intestine;
 RX MEDLINE=93387315; PubMed=8375398;
 RA Agerberth B., Boman A., Andersson M., Joernvall H., Mutt V.,
 Boman H.G.;
 RT "Isolation of three antibacterial peptides from pig intestine:
 RT gastric inhibitory polypeptide (7-42), diazepam-binding inhibitor
 RT (32-86) and a novel factor, peptide 3910."
 RL Eur. J. Biochem. 216:623-629(1993).
 CC -1- FUNCTION: BINDS MEDIUM- AND LONG-CHAIN ACYL-COA ESTERS WITH VERY
 CC HIGH AFFINITY AND MAY FUNCTION AS AN INTRACELLULAR CARRIER OF
 CC ACYL-COA ESTERS. IT IS ALSO ABLE TO DISPLACE DIAZEPAM FROM THE
 CC BENZODIAZEPINE (BZD) RECOGNITION SITE LOCATED ON THE GABA TYPE A
 CC RECEPTOR. IT IS THEREFORE POSSIBLE THAT THIS PROTEIN ALSO ACTS
 CC AS A NEUROPEPTIDE TO MODULATE THE ACTION OF THE GABA RECEPTOR.
 CC -1- FUNCTION: DBI (32-86) HAS ANTIBACTERIAL PROPERTIES.
 CC -1- SUBUNIT: Monomer.
 CC -1- SIMILARITY: Belongs to the ACBP family.
 DR PIR: S00805; NZSG.
 DR HSSP: P07107; IACA.
 DR InterPro: IPR000582; Ac_coa_bind_prot.
 DR Pfam: PF00887; ACBP; 1.
 DR PRINTS: PR00689; ACOABINDINGP.

DR ProDom; PD351532; Ac coa_bind_prot; 1.
 DR PROSITE; PS00880; ACBP; 1.
 KW Transport; Lipid-binding; Acetylation; Antibiotic.
 FT PEPTIDE 32 86 DBI(32-86).
 FT MOD_RES 1 ACETYLATION.
 FT SEQUENCE 86 AA; 9765 MW; EBF12871E07A0F7 CRC64;
 Query Match 1.6%; Score 8; DB 1; Length 86;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 308 KGKAKMDA 315
 DB 50 KGKAKMDA 57

RESULT 6
 ACBP_RANRI STANDARD; PRT; 87 AA.
 AC P45883;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Acyl-CoA-binding protein homolog (ACBP) (Diazepam binding inhibitor homolog) (DBI).
 OS Rana ridibunda (Laughing frog) (Marsh frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Ranidae; Rana.
 NC NCBI_TaxID=8406;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-39 AND 58-87.
 RC TISSUE=Brain;
 RX MEDLINE=94316605; PubMed=8041717;
 RA Lihmann I., Plaquevent J.-C., Tostivint H., Rajmakers R., Tonon M.-C., Conlon J.M., Vaudry H.,
 RT "Frog diazepam-binding inhibitor: peptide sequence, cDNA cloning, and expression in the brain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6899-6903 (1994).
 RT [2]
 RP CHARACTERIZATION.
 RC TISSUE=Brain;
 RX MEDLINE=92319720; PubMed=1341880;
 RA Malagon M., Vaudry H., Vallarino M., Gracia-Navarro F., Tonon M.C.;
 RT "Distribution and characterization of endopeptin-like immunoreactivity in the central nervous system of the frog Rana ridibunda.";
 RL Peptides 13:99-107 (1992).
 CC -1- FUNCTION: MAY PLAY IMPORTANT FUNCTIONS IN THE CONTROL OF BRAIN AND PITUITARY ACTIVITIES. MAY REGULATE GABA NEUROTRANSMISSION THROUGH A PARACRINE AND/OR AUTOCRINE MECHANISM. MAY NOT BIND ACYL-COA ESTERS.
 CC -1- TISSUE SPECIFICITY: BRAIN. IS SELECTIVELY EXPRESSED IN GLIAL CELLS.
 CC -1- SIMILARITY: BELONGS TO THE ACBP FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; U09205; AAB60606.1; -
 DR PIR; A57711; A57711.
 DR HSSP; P07107; IACA.
 DR InterPro; IPR000582; Ac_coa_bind_prot.
 DR Pfam; PF00887; ACBP; 1.
 DR PROSITE; PS00689; ACOABINDINGP.
 DR ProDom; PD351532; Ac_coa_bind_prot; 1.
 DR PROSITE; PS00880; ACBP; 1.
 KW Transport; Lipid-binding.
 FT INIT_MET 0 0

FT PEPTIDE 18 51 TRIKONTATETRANEUROPEPTIDE (TTN)
 FT POTENTIAL.
 FT OCTADECANEUROPEPTIDE (ONN) (POTENTIAL).
 FT SEQUENCE 34 51
 FT MOD_RES 34 51
 FT SEQUENCE 87 AA; 9677 MW; P6621EFC412054BA CRC64;
 Query Match 1.6%; Score 8; DB 1; Length 87;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 308 KGKAKMDA 315
 DB 51 KGKAKMDA 58

RESULT 7
 HSL_SULTO STANDARD; PRT; 283 AA.
 ID HSL_SULTO
 AC Q97023;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ATP phosphoribosyltransferase (EC 2.4.2.17).
 GN HSG OR S71459.
 OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 CC Sulfolobus.
 NC NCBI_TaxID=111955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JCM 10545 / ?;
 RX MEDLINE=21456156; PubMed=11572479;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka R., Kosugi H., Hoshoyama A., Fukui S., Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y., Yoshizawa T., Tanaka T., Kudon Y., Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A., Oshima T., Kikuchi H.;
 RT "Complete genome sequence of an aerobic thermocidophilic Crenarchaeon, Sulfolobus tokodaii strain7.";
 RL DNA Res. 8:123-140 (2001).
 CC -1- CATALYTIC ACTIVITY: 1-(5-phospho-D-ribose)-ATP + diphosphate = ATP + 5-phospho-alpha-D-ribose 1-diphosphate.
 CC -1- PATHWAY: Histidine biosynthesis; first step. Very important in the regulation of histidine metabolism.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE ATP PHOSPHORIBOSYLTRANSFERASE FAMILY. LONG SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; AP000986; BAB6530.1; -
 DR HAMAP; MF_00079; -; 1.
 DR InterPro; IPR001348; ATP_phospho_trans.
 DR Pfam; PF01634; HSG; 1.
 DR ProDom; PD003516; ATP_phospho_trans; 1.
 DR TIGRFAMs; TIGR00070; HSG; 1.
 DR PROSITE; PS01316; ATP_P_PHORIBOSYLTR; 1.
 KW Histidine biosynthesis; Transferase; Glycosyltransferase; Complete proteome.
 KW SEQUENCE 283 AA; 30985 MW; 19C98FC028F73C8C CRC64;
 Query Match 1.6%; Score 8; DB 1; Length 283;
 Best Local Similarity 100.0%; Pred. No. 4.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 404 SPTLSKLA 411
 |||||||

Db 235 SPTUSKLA 242

RESULT 8:

ID	SYBL_ARATH	STANDARD:	PRT:	310 AA.
AC	P59277; Q9C812;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Synlaxin 81 (Atsyp81).			
GN	SYBL1 OR AT1G51740 OR F19C24.5.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	SpERMATOPHYTES; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	euroside II; Brassicales; Brassicaceae; Arabidopsia.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RX	MEDLINE=21016719; PubMed=11130712;			
RA	Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,			
RA	White O., Alonso J., Alcafi H., Araujo R., Bowman C.L., Brooks S.Y.,			
RA	Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,			
RA	Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,			
RA	Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,			
RA	Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,			
RA	Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,			
RA	Kim C.J., Koo H.L., Kremenetskaia I., Kutz D.B., Kwan A., Lam B.,			
RA	Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,			
RA	Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maltz R., Marshall A.,			
RA	Miltchev J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,			
RA	Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,			
RA	Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,			
RA	Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,			
RA	Utebbeck T., Van Aken S., Vayberg M., Vysotskaia V.S., Walker M.,			
RA	Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;			
RT	"Sequence and analysis of chromosome 1 of the plant Arabidopsis			
RT	thaliana.";			
RL	Nature 408:816-820(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Bryover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,			
RA	Feldmann K.;			
RT	"Full-length cDNA from Arabidopsis thaliana.";			
RL	Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.			
CC	-1- FUNCTION: Vesicle trafficking protein that functions in the			
CC	secretory pathway (By similarity).			
CC	-1- SUBUNIT: Part of the t-SNARE complex (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Type IV membrane protein (By similarity).			
CC	-1- SIMILARITY: BELONGS TO THE SYNTAXIN/EPIMORPHIN FAMILY.			
CC	-1- CAUTION: Ref.1 sequence differs from that shown due to erroneous			
CC	gene model prediction.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL AC025294; AAG50888.1; ALT_SEQ.			
DR	EMBL AY086559; AAM63632.1;			
DR	PROSITE PS00914; SYNTAXIN; FALSE_NEG.			
KW	Transport; Protein transport; Transmembrane; Coiled coil;			
KW	Multi-gene family.			
FT	DOMAIN 1 289			
FT	TRANSMEM 290 310			
FT	ANCHOR FOR TYPE IV MEMBRANE PROTEIN			
FT	(POTENTIAL).			
FT	COILED COIL (POTENTIAL).			
FT	SEQUENCE 310 AA; 35587 MW; 992C2B5F5B87B60 CRC64;			

Query Match 1.6%; Score 8; DB 1; Length 310;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 LGNKELSQ 363

Db 273 LGNKELSQ 280

RESULT 9

ID	WR21_ARATH	STANDARD:	PRT:	380 AA.
AC	O04336;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Probable WRKY transcription factor 21 (WRKY DNA-binding protein 21).			
GN	WRKY21 OR AT2G30590 OR T6B20.6 OR T6B20.6.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	SpERMATOPHYTES; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	euroside II; Brassicales; Brassicaceae; Arabidopsia.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A. AND INDUCTION.			
RC	MEDLINE=21342502; PubMed=11449049;			
RA	Yu D., Chen C., Chen Z.;			
RT	"Evidence for an important role of WRKY DNA binding proteins in the			
RT	regulation of NPR1 gene expression.";			
RL	Plant Cell 13:1527-1540(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RX	MEDLINE=20083487; PubMed=10617197;			
RA	Lin X., Kaul S., Roundley S.D., Shea T.P., Benito M.-I., Town C.D.,			
RA	Fujii C.Y., Mason T.M., Bowman C.L., Barneshead M.E., Feldblum T.V.,			
RA	Buell C.Y., Ketchum K.A., Lee J.D., Ronning C.M., Koo H.L.,			
RA	Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,			
RA	Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,			
RA	Goodman H.M., Somerville C.R., Copenhaver G.P., Freese D.,			
RA	Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,			
RT	Venter J.C.;			
RT	"Sequence and analysis of chromosome 2 of the plant Arabidopsis			
RT	thaliana.";			
RL	Nature 402:761-768(1999).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Columbia;			
RA	Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;			
RT	"Riken Arabidopsis full length cDNA clones (RAFLs) sequenced by the			
RT	SSP consortium (Salk/Stanford/PGEC).";			
RL	Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.			
CC	-1- FUNCTION: Transcription factor. Interacts specifically with the W			
CC	box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-			
CC	responsive cis-acting element (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Nuclear (Probable).			
CC	-1- INDUCTION: By salicylic acid.			
CC	-1- SIMILARITY: Belongs to the WRKY group II-d family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL AF272747; AAK28441.1;			
DR	EMBL U93215; AAB63078.1;			
DR	EMBL BT000231; AAN15550.1;			
DR	EMBL AY136441; AAM97106.1;			
DR	PIR; C84710; C84710.			

```

DR InterPro: IPR003657; WRKY.
DR Pfam: PF03106; WRKY. 1.
DR PROSITE: PSS0811; WRKY. 1.
KM Transcription regulation; DNA-binding; Nuclear protein.
FT DNA_BIND 307 373 WRKY.
FT DOMAIN 154 178 POLY-GLN.
FT DOMAIN 190 195 POLY-GLN.
SQ SEQUENCE 380 AA; 43005 MW; AEE49BFD6EFD797 CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 1; Length 380;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 SSGFQELSL 28
Db 103 SSGFQELSL 110

RESULT 10
YMF8 YEAST STANDARD; PRT; 504 AA.
AC 004931;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypochemical 56.2 kDa protein in ERG8-UBP4 intergenic region.
GN YMR221C OR YMR959.03C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII."
RL Nature 387:90-93(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z49939; CAA00192.1; -.
DR PIR: S57589; S57589.
DR SGD: S0004834; YMR221C.
KM Hypochemical protein; Transmembrane.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 65 85 POTENTIAL.
FT TRANSMEM 92 112 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 187 207 POTENTIAL.
FT TRANSMEM 303 323 POTENTIAL.
FT TRANSMEM 345 365 POTENTIAL.
FT TRANSMEM 386 406 POTENTIAL.
FT TRANSMEM 422 442 POTENTIAL.
FT TRANSMEM 463 483 POTENTIAL.
SQ SEQUENCE 504 AA; 56177 MW; 38C5E75DFB874767 CRC64;

Query Match
Best Local Similarity 1.6%; Score 8; DB 1; Length 504;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 240 YLVGYTL 247
Db 121 YLVGYTL 128

RESULT 11
CDPK SOYBN STANDARD; PRT; 508 AA.
ID P28583;
AC 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Calcium-dependent protein kinase SK5 (EC 2.7.1.-) (CDPK).
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucoside I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Williams;
RX MEDLINE=91240279; PubMed=1852075;
RA Harper J.F., Sussman M.R., Schaller G.E., Putnam-Evans C.,
RA Chabonneau H., Harmon A.C.;
RT "A calcium-dependent protein kinase with a regulatory domain similar
RT to calmodulin."
RL Science 252:951-954(1991).
CC -1- FUNCTION: MAY PLAY A ROLE IN SIGNAL TRANSDUCTION PATHWAYS THAT
CC INVOLVE CALCIUM AS A SECOND MESSENGER.
CC -1- ENZYME REGULATION: ACTIVATED BY CALCIUM. AUTOPHOSPHORYLATION MAY
CC PLAY AN IMPORTANT ROLE IN THE REGULATION OF THE KINASE ACTIVITY.
CC -1- TISSUE SPECIFICITY: FOUND THROUGHOUT THE PLANT.
CC -1- MISCELLANEOUS: THERE ARE MULTIPLE CDPK ISOFORMS IN SOYBEAN.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CAK SUBFAMILY.
CC -1- SIMILARITY: Contains 4 EF-hand calcium-binding domains.
CC -1- SIMILARITY: Contains 4 EF-hand calcium-binding domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M64987; AAB0806.1; -.
DR PIR: A43713; A43713.
DR HSSP: O63450; 1A06.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000719; Prot kinase.
DR InterPro: IPR002290; Ser_Thr_kinase.
DR Pfam: PF00036; ehand; 4.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000012; EF-hand; 2.
DR ProDom: PD000001; Prot_kinase; 1.
DR SMART: SM00054; Eph; 4.
DR SMART: SM00220; S_TKc; 1.
DR PROSITE: PSS00018; EF_HAND; 4.
DR PROSITE: PSS0107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PSS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PSS0011; PROTEIN_KINASE_DOM; 1.
KM Transferrase; Serine/threonine-protein kinase; ATP-binding;
KM Calcium-binding; Phosphorylation.
FT DOMAIN 34 292 PROTEIN KINASE.
FT NP_BIND 40 48 ATP (BY SIMILARITY).
FT BINDING 63 63 ATP (BY SIMILARITY).
FT ACT_SITE 158 158 BY SIMILARITY.
FT CA_BIND 348 359 EF_HAND 1 (POTENTIAL).
FT CA_BIND 384 395 EF_HAND 2 (POTENTIAL).
FT CA_BIND 420 431 EF_HAND 3 (POTENTIAL).
FT CA_BIND 454 465 EF_HAND 4 (POTENTIAL).
SQ SEQUENCE 508 AA; 57169 MW; AFD5C5122419284 CRC64;

```

Query Match 1.6%; Score 8; DB 1; Length 508;
 Best Local Similarity 100.0%; Pred. No. 8.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 394 EROAARLI 401
 |||||
 DB 132 EROAARLI 139

RESULT 12

AMP2_MELGA
 ID AMP2_MELGA STANDARD; PRT; 64 AA.
 AC P80392; O93506;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DE 15-JUL-1999 (Rel. 38, Last annotation update)
 OS Antimicrobial peptide THP2 precursor (Turkey heterophil peptide 2).
 OC Melasaris gallipavo (Common turkey).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Meleagrididae; Meleagris.
 NCBI_TaxID=9103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=98418188; PubMed=9745666;
 RA Brooks C.W., Harmon B.G., Jackwood M.W.;
 RT "Characterization of beta-defensin prepropeptide mRNA from chicken
 and turkey bone marrow."
 RL Anim. Genet. 29:283-289(1998).
 RN [2]
 RP SEQUENCE OF 29-48.
 RX MEDLINE=95053386; PubMed=7964174;
 RA Evans E.W., Beach G.G., Wunderlich J., Harmon B.G.;
 RT "Isolation of antimicrobial peptides from avian heterophils."
 RL J. Leukoc. Biol. 56:661-665(1994).
 CC -1- SIMILARITY: BELONGS TO THE BETA-DEFENSIN FAMILY.
 CC -1- SIMILARITY: BELONGS TO THE BETA-DEFENSIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF033338; AAC36054.1; -
 CC
 KW Antibiotic; Signal.
 FT SIGNAL 1 28
 FT CHAIN 29 64 ANTIMICROBIAL PEPTIDE THP2.
 FT DISULFID 31 57 BY SIMILARITY.
 FT DISULFID 36 51 BY SIMILARITY.
 FT DISULFID 41 58 BY SIMILARITY.
 SQ SEQUENCE 64 AA; 7327 MW; 7E597FEE4FCA4995 CRC64;
 Query Match 1.4%; Score 7; DB 1; Length 64;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Meleagrididae;
 CC Gallus.
 NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=98418188; PubMed=9745666;
 RA Brooks C.W., Harmon B.G., Jackwood M.W.;
 RT "Characterization of beta-defensin prepropeptide mRNA from chicken
 and turkey bone marrow."
 RL Anim. Genet. 29:283-289(1998).
 RN [2]

RP SEQUENCE OF 29-64, AND CHARACTERIZATION.
 RC STRAIN=Cross Broiler-6; TISSUE=Leukocyte;
 RX MEDLINE=94200386; PubMed=8150085;
 RA Harwig S.S.L., Swider K.M., Kokryakov V.N., Tan L., Lee T.D.,
 RA Panyutich E.A., Aleshina G.M., Shamova O.V., Lehrer R.I.;
 RT "Gallinacins: cysteine-rich antimicrobial peptides of chicken
 leukocytes."
 RL FEBS Lett. 342:281-285(1994).
 CC -1- FUNCTION: HAS BACTERICIDAL ACTIVITY. POTENT AGAINST E. COLI ML-35,
 AND L. MONOCYTOGENES EGD BUT NOT C. ALBICANS.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic granules.
 CC -1- SIMILARITY: BELONGS TO THE BETA-DEFENSIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF033336; AAC36052.1; -
 CC
 KW Antibiotic; Signal.
 FT SIGNAL 1 28
 FT CHAIN 29 64 GALLINACIN 2.
 FT DISULFID 31 57 BY SIMILARITY.
 FT DISULFID 36 51 BY SIMILARITY.
 FT DISULFID 41 58 BY SIMILARITY.
 SQ SEQUENCE 64 AA; 7141 MW; 73E147BE4FCD39E0 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 64;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 366 SLFLAL 372
 |||||
 DB 9 SLFLAL 15

RESULT 14

ACBP CANFA
 ID ACBP CANFA STANDARD; PRT; 86 AA.
 AC O970X6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 OS Acyl-CoA-binding protein (ACBP) (Diazepam binding inhibitor) (DBI)
 DE (Endozepine) (EP).
 GN DBI.
 OS Canis familiaris (Dog).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=9619960; PubMed=8609609;
 RA Kragelund B.B., Hoeyrup P., Jensen M.S., Schjerling C.K., Juul E.,
 RA Knudsen J., Poulsen F.M.;
 RT "Fast and one-step folding of closely and distantly related homologous
 proteins of a four-helix bundle family."

RL J. Mol. Biol. 256:187-200(1996).
 CC -1- FUNCTION: BINDS MEDIUM- AND LONG-CHAIN ACYL-COA ESTERS WITH VERY
 CC HIGH AFFINITY AND MAY FUNCTION AS AN INTRACELLULAR CARRIER OF
 CC ACYL-COA ESTERS.
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SIMILARITY: Belongs to the ACPB family.
 DR PIR: S63592; S63592.
 DR HSSP: P07107; IACA.
 DR InterPro: IPR000582; Ac_coa_bind_prot.
 DR Pfam: PF00887; ACPB; 1.
 DR PRINTS: PR00689; ACOABINDING.
 DR ProDom: PD51532; Ac_coa_bind_prot; 1.
 DR PROSITE: PS00880; ACPB; 1.
 DR Transport; Lipid-binding; Acetylation.
 FT INIT MET 0
 FT MOD RES 1 1 ACETYLATION (BY SIMILARITY).
 SQ SEQUENCE 86 AA; 9847 MW; 654655207E9B59D CRC64;
 QY Query Match 1.4%; Score 7; DB 1; Length 86;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 309 GSAKMDA 315
 51 GSAKMDA 57
 RESULT 15
 ACPB_HUMAN STANDARD; PRT; 86 AA.
 ID ACPB_HUMAN STANDARD; PRT; 86 AA.
 AC P07108; P08869;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-NOV-1988 (Rel. 09, Last annotation update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Acyl-CoA-binding protein (ACBP) (Diazepam binding inhibitor) (DBI)
 DE (Benzodiazepine) (BZ).
 GN DBI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RC TISSUE=Brain;
 RP MEDLINE=86278003; PubMed=3525533;
 RA Marguier H., Todaro G.J., Shoyab M.;
 RT "Complete amino acid sequences of bovine and human endopeptidases.
 RT Homology with rat diazepam binding inhibitor.";
 RL J. Biol. Chem. 261:9727-9731(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87161236; PubMed=2881742;
 RA Webb N.R., Rose T.M., Malik N., Marguier H., Shoyab M., Todaro G.J.,
 RA Lee D.C.;
 RT "Bovine and human cDNA sequences encoding a putative benzodiazepine
 RT receptor ligand.";
 RL DNA 6:71-79(1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87016986; PubMed=3020548;
 RA Gray P.W., Gialister D., Seeburg P.H., Guidotti A., Costa E.;
 RT "Cloning and expression of cDNA for human diazepam binding inhibitor,
 RT a natural ligand of an allosteric regulatory site of the gamma-
 RT aminobutyric acid type A receptor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:7547-7551(1986).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Brain;
 RX MEDLINE=2338257; PubMed=12477932;
 RA Strussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schler G.D.,
 RA Altschul S.F., Zeeberg B., Buecaw K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carinini P., Prange C.,
 RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McKernan K.J., Malek A.M., Gay L.J., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Scherich A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [5]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=95194304; PubMed=7534063;
 RA Kolner M., Rovio A., Alho H.;
 RT "The characterization of two diazepam binding inhibitor (DBI)
 RT transcripts in humans.";
 RL Biochem. J. 306:327-330(1995).
 CC -1- FUNCTION: BINDS MEDIUM- AND LONG-CHAIN ACYL-COA ESTERS WITH VERY
 CC HIGH AFFINITY AND MAY FUNCTION AS AN INTRACELLULAR CARRIER OF
 CC ACYL-COA ESTERS. IT IS ALSO ABLE TO DISPLACE DIAZEPAM FROM THE
 CC BENZODIAZEPINE (BZD) RECOGNITION SITE LOCATED ON THE GABA TYPE A
 CC RECEPTOR. IT IS THEREFORE POSSIBLE THAT THIS PROTEIN ALSO ACTS AS
 CC A NEUROPEPTIDE TO MODULATE THE ACTION OF THE GABA RECEPTOR.
 CC -1- SUBUNIT: Monomer.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=Short;
 CC IsoId=P07108-1; Sequence=Displayed.
 CC Name=2; Synonyms=Long;
 CC IsoId=P07108-2; Sequence=VSP 000068;
 CC -1- SIMILARITY: Belongs to the ACPB family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@ib-sib.ch).
 CC -----
 DR EMBL: M14200; AAAS2171.1; -;
 DR EMBL: M15887; AAA35788.1; -;
 DR EMBL: BC006466; AAH06466.1; -;
 DR PIR: B26448; NZHU.
 DR HSSP: P07107; IACA.
 DR Genew: HGNC:2690; DBI.
 DR MIM: 125950;
 DR GO: GO:0008025; F:diazepam-binding inhibitor activity; TAS.
 DR GO: GO:0005102; F:receptor binding activity; TAS.
 DR InterPro: IPR000582; Ac_coa_bind_prot.
 DR Pfam: PF00887; ACPB; 1.
 DR PRINTS: PR00689; ACOABINDING.
 DR ProDom: PD51532; Ac_coa_bind_prot; 1.
 DR PROSITE: PS00880; ACPB; 1.
 DR Transport; Lipid-binding; Acetylation; Alternative splicing.
 FT INIT MET 0
 FT MOD RES 1 1 ACETYLATION.
 FT VARSPLOC 1 2 SO -> MNGDWLPPASNPOTGTE (in isoform
 FT 2).
 SQ SEQUENCE 86 AA; 9913 MW; 6A973E406CB1D0E7 CRC64;
 QY Query Match 1.4%; Score 7; DB 1; Length 86;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 309 GSAKMDA 315

Db |||
 51 GKAKMDA 57

Search completed: November 21, 2003, 14:23:07
Job time : 19 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 14:21:44 ; Search time 21 Seconds
(without alignments)
981.209 Million cell updates/sec

Title: US-09-938-803-6

Perfect score: 487
Sequence: 1 MASSAEDEGTVALAGVLQ.....ILPYIKAMVPRKELKKME 487

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	1.6	86	1 US-08-700-626-5	Sequence 5, Appl1
2	8	1.6	463	1 US-07-951-715A-25	Sequence 25, Appl1
3	8	1.6	463	2 US-08-459-448A-25	Sequence 25, Appl1
4	8	1.6	463	3 US-08-459-595A-25	Sequence 25, Appl1
5	8	1.6	463	3 US-08-459-504B-25	Sequence 25, Appl1
6	8	1.6	463	3 US-08-459-444-25	Sequence 0, Appl1
7	8	1.6	463	4 US-09-547-442-25	Sequence 0, Appl1
8	8	1.6	947	4 US-09-228-986-73	Sequence 73, Appl1
9	7	1.4	28	1 US-08-182-175A-4	Sequence 4, Appl1
10	7	1.4	28	1 US-08-182-175A-49	Sequence 49, Appl1
11	7	1.4	28	1 US-08-182-175A-83	Sequence 83, Appl1
12	7	1.4	28	1 US-08-474-633A-58	Sequence 58, Appl1
13	7	1.4	28	1 US-08-474-633A-73	Sequence 58, Appl1
14	7	1.4	28	4 US-08-823-771-58	Sequence 58, Appl1
15	7	1.4	28	4 US-08-823-771-73	Sequence 73, Appl1
16	7	1.4	28	4 US-09-351-657A-34	Sequence 34, Appl1
17	7	1.4	28	4 US-09-351-657A-35	Sequence 35, Appl1
18	7	1.4	28	5 PCT-US92-06412-4	Sequence 4, Appl1
19	7	1.4	28	5 PCT-US92-06412-49	Sequence 49, Appl1
20	7	1.4	28	5 PCT-US92-06412-83	Sequence 83, Appl1
21	7	1.4	31	1 US-08-190-802A-219	Sequence 219, Appl1
22	7	1.4	31	3 US-08-477-346-219	Sequence 219, Appl1
23	7	1.4	31	4 US-08-477-346-219	Sequence 219, Appl1
24	7	1.4	31	4 US-08-487-072A-219	Sequence 219, Appl1
25	7	1.4	40	1 US-08-144-121-10	Sequence 10, Appl1
26	7	1.4	40	2 US-08-735-893-10	Sequence 10, Appl1
27	7	1.4	60	4 US-09-149-476-403	Sequence 403, Appl1

28	7	1.4	64	4 US-09-351-657A-4	Sequence 4, Appl1
29	7	1.4	64	4 US-09-351-657A-8	Sequence 8, Appl1
30	7	1.4	77	1 US-08-182-175A-57	Sequence 57, Appl1
31	7	1.4	77	1 US-08-474-633A-75	Sequence 75, Appl1
32	7	1.4	77	4 US-08-823-771-75	Sequence 75, Appl1
33	7	1.4	77	5 PCT-US92-06412-57	Sequence 57, Appl1
34	7	1.4	86	1 US-08-700-626-4	Sequence 4, Appl1
35	7	1.4	86	4 US-09-300-008B-55	Sequence 55, Appl1
36	7	1.4	104	1 US-08-700-626-3	Sequence 3, Appl1
37	7	1.4	187	4 US-09-252-991A-29325	Sequence 29325, A
38	7	1.4	188	2 US-08-484-993B-20	Sequence 20, Appl1
39	7	1.4	188	2 US-08-484-158B-20	Sequence 20, Appl1
40	7	1.4	188	2 US-08-484-596A-20	Sequence 20, Appl1
41	7	1.4	188	2 US-08-480-150A-20	Sequence 20, Appl1
42	7	1.4	188	3 US-08-458-731-20	Sequence 20, Appl1
43	7	1.4	188	3 US-08-149-223A-20	Sequence 20, Appl1
44	7	1.4	203	4 US-09-134-001C-3145	Sequence 3145, Ap
45	7	1.4	257	4 US-09-252-991A-23705	Sequence 23705, A

ALIGNMENTS

RESULT 1
US-08-700-626-5
; Sequence 5, Application US/08700626
; Patent No. 5734038
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN DBI/ACBP-LIKE PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700,626
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0115 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 86 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 118275
; US-08-700-626-5
Query Match 1.6%; Score 8; DB 1; Length 86;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
QY 308 GKAKWDA 315
|||||||

DB 50 KOKAKMDA 57

RESULT 2

US-07-951-715A-25
Sequence 25, Application US/07951715A

Patent No. 5625136

GENERAL INFORMATION:

APPLICANT: Koziet, Michael G.
APPLICANT: Desai, Nalini M.
APPLICANT: Lewis, Kelly S.
APPLICANT: Kramer, Vance C.
APPLICANT: Warren, Gregory W.
APPLICANT: Evola, Stephen V.
APPLICANT: Crossland, Lyle D.
APPLICANT: Wright, Martha S.
APPLICANT: Merlin, Ellis J.
APPLICANT: Launis, Karen L.
APPLICANT: Rothstein, Steven J.
APPLICANT: Bowman, Cindy G.
APPLICANT: Dawson, John L.
APPLICANT: Dunder, Erik M.
APPLICANT: Pace, Gary M.
APPLICANT: Suttie, Janet L.
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
INSECTICIDAL ACTIVITY IN MAIZE
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/951, 715A
FILING DATE: 25-SEP-1992
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/772, 027
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: S-18805/A/CGC 1577/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8615
TELEFAX: (919)541-8689

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: Protein
LOCATION: 1..463
OTHER INFORMATION: /note= "protein sequence for
soybean CDPK as shown in Figure 34."
US-07-951-715A-25

Query Match 1.6%; Score 8; DB 1; Length 463;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 394 ERQARLI 401

DB 111 ERQARLI 118

RESULT 3

US-08-459-448A-25
Sequence 25, Application US/08459448A

Patent No. 5659336

GENERAL INFORMATION:

APPLICANT: Koziet, Michael G.
APPLICANT: Desai, Nalini M.
APPLICANT: Lewis, Kelly S.
APPLICANT: Kramer, Vance C.
APPLICANT: Warren, Gregory W.
APPLICANT: Evola, Stephen V.
APPLICANT: Crossland, Lyle D.
APPLICANT: Wright, Martha S.
APPLICANT: Merlin, Ellis J.
APPLICANT: Launis, Karen L.
APPLICANT: Rothstein, Steven J.
APPLICANT: Bowman, Cindy G.
APPLICANT: Dawson, John L.
APPLICANT: Dunder, Erik M.
APPLICANT: Pace, Gary M.
APPLICANT: Suttie, Janet L.
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
INSECTICIDAL ACTIVITY IN MAIZE
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5659336artis Corporation
STREET: Patent & Trademark Dept., 520 White Plains
STREET: Rd., POB 2005
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-9005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459, 448A
FILING DATE: 02-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/951, 715
FILING DATE: 25-SEP-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/772, 027
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40403
REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8582
TELEFAX: (919)541-8689

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: Protein
LOCATION: 1..463
OTHER INFORMATION: /note= "protein sequence for
soybean CDPK as shown in Figure 34."
US-08-459-448A-25

Query Match 1.6%; Score 8; DB 1; Length 463;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 1.6%; Score 8; DB 2; Length 463;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 394 EROARLI 401
DB 111 EROARLI 118

RESULT 4
US-08-459-595A-25
Sequence 25, Application US/08459595A
Patent No. 6018104

GENERAL INFORMATION:
APPLICANT: Koziel, Michael G.
APPLICANT: Desai, Nalini M.
APPLICANT: Lewis, Kelly S.
APPLICANT: Kramer, Vance C.
APPLICANT: Warren, Gregory W.
APPLICANT: Evola, Stephen V.
APPLICANT: Crossland, Lyle D.
APPLICANT: Wright, Martha S.
APPLICANT: Merlin, Ellis J.
APPLICANT: Launis, Karen L.
APPLICANT: Rothstein, Steven J.
APPLICANT: Bowman, Cindy G.
APPLICANT: Dawson, John L.
APPLICANT: Dunder, Erik M.
APPLICANT: Pace, Gary M.
APPLICANT: Suttie, Janet L.
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 601810artis Corporation
STREET: Patent & Trademark Dept., 520 White Plains
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-9005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,595A
FILING DATE: 02-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/951,715
FILING DATE: 25-SEP-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/772,027
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40403
REFERENCE/DOCKET NUMBER: CGC 1577/CIP/DIV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8582
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: Protein

LOCATION: 1..463 /note= "protein sequence for
OTHER INFORMATION: soybean CDPK as shown in Figure 34."
US-08-459-595A-25

Query Match 1.6%; Score 8; DB 3; Length 463;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 394 EROARLI 401
DB 111 EROARLI 118

RESULT 5
US-08-459-504B-25
Sequence 25, Application US/08459504B
Patent No. 6075185

GENERAL INFORMATION:
APPLICANT: Koziel, Michael G.
APPLICANT: Desai, Nalini M.
APPLICANT: Lewis, Kelly S.
APPLICANT: Kramer, Vance C.
APPLICANT: Warren, Gregory W.
APPLICANT: Evola, Stephen V.
APPLICANT: Crossland, Lyle D.
APPLICANT: Wright, Martha S.
APPLICANT: Merlin, Ellis J.
APPLICANT: Launis, Karen L.
APPLICANT: Rothstein, Steven J.
APPLICANT: Bowman, Cindy G.
APPLICANT: Dawson, John L.
APPLICANT: Dunder, Erik M.
APPLICANT: Pace, Gary M.
APPLICANT: Suttie, Janet L.
TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
TITLE OF INVENTION: INSECTICIDAL ACTIVITY IN MAIZE
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 607518artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,504B
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/459,595
FILING DATE: 02-JUN-1995
APPLICATION NUMBER: US 07/951,715
FILING DATE: 25-SEP-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/772,027
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1577/CIP/DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8587
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: Protein
LOCATION: 1..463
OTHER INFORMATION: /note= "protein sequence for
US-08-459-504B-25 soybean CDPK as shown in Figure 34."

Query Match 1.6%; Score 8; DB 3; Length 463;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 394 EROARLI 401
DB 111 EROARLI 118

RESULT 6
US-08-459-444-25
Sequence 0, Application US/08459444A
Patent No. 6121014
GENERAL INFORMATION:
APPLICANT: Kozziel, Michael G.
Desai, Nalini M.
Lewis, Kelly S.
Kramer, Vance C.
Warren, Gregory W.
Evola, Stephen V.
Crossland, Lyle D.
Wright, Martha S.
Merlin, Ellis J.
Lauais, Karen L.

TITLE OF INVENTION: METHOD FOR PRODUCING A PLANT-OPTIMIZED
NUCLEIC ACID CODING SEQUENCE

NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6121014artis Agribusiness Biotechnology Research, Inc.
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,444A
FILING DATE: 02-Jun-1995
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/951,715
FILING DATE: 25-SEP-1992
APPLICATION NUMBER: US 07/772,027
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-18805/PI/CGC1577/CIP/DIV6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8587
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: /note= "protein sequence for
soybean CDPK as shown in Figure 34."

SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: Protein
LOCATION: 1..463
SEQUENCE DESCRIPTION: SEQ ID NO: 25:
US-08-459-444-25

Query Match 1.6%; Score 8; DB 3; Length 463;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 394 EROARLI 401
DB 111 EROARLI 118

RESULT 7
US-09-547-422-25
Sequence 0, Application US/09547422
Patent No. 6320100
GENERAL INFORMATION:
APPLICANT: Kozziel, Michael G.
Desai, Nalini M.
Lewis, Kelly S.
Kramer, Vance C.
Warren, Gregory W.
Evola, Stephen V.
Crossland, Lyle D.
Wright, Martha S.
Merlin, Ellis J.
Lauais, Karen L.

TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
INSECTICIDAL ACTIVITY IN MAIZE

NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6320100artis Agribusiness Biotechnology Research, Inc.
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/547,422
FILING DATE: 11-Apr-2000
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/459,595
FILING DATE: 02-JUN-1995
APPLICATION NUMBER: US 07/951,715
FILING DATE: 25-SEP-1992
APPLICATION NUMBER: US 07/772,027
FILING DATE: 04-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-18805H
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8587
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: /note= "protein sequence for
soybean CDPK as shown in Figure 34."

SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

HYPOTHETICAL: NO
FEATURE:
NAME/KEY: Protein
LOCATION: 1..463
SEQUENCE DESCRIPTION: SEQ ID NO: 25:
US-09-547-422-25

Query Match 1.6%; Score 8; DB 4; Length 463;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 394 EROARLI 401
Db 111 EROARLI 118

RESULT 8
US-09-228-986-73
Sequence 73, Application US/09228986
Patent No. 6359198
GENERAL INFORMATION:
APPLICANT: Strabala, Timothy
APPLICANT: Nieuwenhuizen, Niels
TITLE OF INVENTION: Compositions Isolated from Plant Cells
TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
FILE REFERENCE: 11000/1020
CURRENT APPLICATION NUMBER: US/09/228,986
CURRENT FILING DATE: 1999-01-12
NUMBER OF SEQ ID NOS: 130
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 73
LENGTH: 947
TYPE: PRT
ORGANISM: Pinus radiata
US-09-228-986-73

Query Match 1.6%; Score 8; DB 4; Length 947;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 462 FLSLFIL 469
Db 10 FLSLFIL 17

RESULT 9
US-08-182-175A-4
Sequence 4, Application US/08182175A
Patent No. 5559223
GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
APPLICANT: Janet A. Rice
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/182,175A
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006

FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:
NAME: Linda Axamethy Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 28 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..28
OTHER INFORMATION: /label= name
US-08-182-175A-4
OTHER INFORMATION: /note= "(SSP 8)4"

Query Match 1.4%; Score 7; DB 1; Length 28;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 481 EKLKME 487
Db 3 EKLKME 9

RESULT 10
US-08-182-175A-49
Sequence 49, Application US/08182175A
Patent No. 5559223
GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
APPLICANT: Janet A. Rice
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing P
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/182,175A
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:
NAME: Linda Axamethy Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 28 amino acids
TYPE: amino acid

```

;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-08-182-175A-49

```

Query Match	1.4%	Score 7;	DB 1;	Length 28;
Best Local Similarity	100.0%	Pred. No. 13;		
Matches	7;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0

QY	481	EKLKME	487
Db	10	EKLKME	16

RESULT 11
US-08-182-175A-83

APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
APPLICANT: Janet A. Rice
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing I
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;

```

Query Match	1.4%	Score 7	DB 1	Length 28
Best Local	Similarity 100.0%	Pred. NO. 13		
Matches 7	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	481	EKLKME	487
Db	3	EKLKME	9

RESULT 12
US-08-474-633A-58
; Sequence 58, Application US/08474633A
; Patent No. 5773691
; GENERAL INFORMATION:
;

1 APPLICANT: E. I. DU PONT DE NEMOURS AND
2 APPLICANT: COMPANY
3 TITLE OF INVENTION: CHIMERIC GENES AND
4 TITLE OF INVENTION: METHODS FOR INCREASING
5 TITLE OF INVENTION: INCREASING THE LYSINE
6 TITLE OF INVENTION: AND THREONINE CONTENT
7 TITLE OF INVENTION: OF THE SEEDS OF PLANTS
8 NUMBER OF SEQUENCES: 107
9
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: E. I. DU PONT DE NEMOURS
12
13 ADDRESSEE: AND COMPANY
14 STREET: 1007 MARKET STREET
15 CITY: WILMINGTON
16 STATE: DELAWARE
17 COUNTRY: U.S.A.
18
19 000000

```

1 COMPUTER READABLE FORM:
2 MEDIUM TYPE: FLOPPY DISK
3 COMPUTER: IBM PC COMPATIBLE
4 OPERATING SYSTEM: PC-DOS/MS-DOS
5 SOFTWARE: MICROSOFT WORD VERSION 2.00
6 CURRENT APPLICATION DATA:
7 APPLICATION NUMBER: US/08/474,633A
8 FILING DATE:
9

```

Query Match	1.4%	Score 7	DB 1	Length 28
Best Local Similarity	100.0%	Pred. No. 13		
Matches 7	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	481	EKLKME	487
Db	10	EKLKME	16

RESULT 13
US-08-474-633A-73
; Sequence 73, Application US/08474633A

1 APPLICANT: E. I. DU PONT DE NEMOURS AND
2 APPLICANT: COMPANY
3 TITLE OF INVENTION: CHIMERIC GENES AND
4 TITLE OF INVENTION: METHODS FOR INCREASING
5 TITLE OF INVENTION: INCREASING THE LYSINE
6 TITLE OF INVENTION: AND THREONINE CONTENT
7 TITLE OF INVENTION: OF THE SEEDS OF PLANTS
8 NUMBER OF SEQUENCES: 107
9 CORRESPONDENCE ADDRESS: 107

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
;

```

COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD VERSION 2.0C
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,633A
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: BARBARA C. SIEGELL
REGISTRATION NUMBER: 30,684
REFERENCE/DOCKET NUMBER: BB-1037-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4931
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 28 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-474-633A-73

Query Match 1.4%; Score 7; DB 1; Length 28;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 481 EKLKME 487
Db 3 EKLKME 9

RESULT 14
US-08-823-771-58
Sequence 58, Application US/08823771
Patent No. 6459019
GENERAL INFORMATION:
APPLICANT: E. I. DU PONT DE NEMOURS AND
COMPANY
TITLE OF INVENTION: CHIMERIC GENES AND
METHODS FOR INCREASING
INCREASING THE LYSINE
AND THREONINE CONTENT
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS
AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD VERSION 2.0C
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,771
FILING DATE: 24-Mar-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/474,633
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BARBARA C. SIEGELL
REGISTRATION NUMBER: 30,684
REFERENCE/DOCKET NUMBER: BB-1037-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4931
TELEFAX: 302-773-0164

INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 28 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-08-823-771-58

Query Match 1.4%; Score 7; DB 4; Length 28;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 481 EKLKME 487
Db 10 EKLKME 16

RESULT 15
US-08-823-771-73
Sequence 73, Application US/08823771
Patent No. 6459019
GENERAL INFORMATION:
APPLICANT: E. I. DU PONT DE NEMOURS AND
COMPANY
TITLE OF INVENTION: CHIMERIC GENES AND
METHODS FOR INCREASING
INCREASING THE LYSINE
AND THREONINE CONTENT
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS
AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD VERSION 2.0C
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,771
FILING DATE: 24-Mar-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/474,633
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BARBARA C. SIEGELL
REGISTRATION NUMBER: 30,684
REFERENCE/DOCKET NUMBER: BB-1037-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4931
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 28 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-08-823-771-73

Query Match 1.4%; Score 7; DB 4; Length 28;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 481 EKLKME 487

|||||
Db 3 EXTRACT 9

Search completed: November 21, 2003, 14:25:05
Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 14:24:05 ; Search time 36 Seconds

(without alignments)
2469.624 Million cell updates/sec

Title: US-09-938-803-6

Perfect score: 487
Sequence: 1 MASSAEDEGTVALAGVLA.....ILPYIKAMVPRKELKKKE 487Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 666188 seqs, 182559486 residues

Word size : 0

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Published Applications AA.*

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTOS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	487	100.0	487	9	US-09-938-803-6
2	56	11.5	57	12	US-10-029-386-29742
3	42	8.6	58	9	US-09-864-761-33415
4	8	1.6	86	9	US-09-859-101-5
5	8	1.6	86	10	US-09-987-108-1
6	8	1.6	86	10	US-09-987-108-3
7	8	1.6	86	10	US-09-987-108-7
8	8	1.6	86	10	US-09-987-108-8
9	8	1.6	86	10	US-09-987-108-9
10	8	1.6	86	10	US-09-987-108-20
11	8	1.6	463	11	US-09-988-462-25
12	8	1.6	609	12	US-10-032-585-7294
13	8	1.6	947	15	US-10-101-464A-73
14	7	1.4	20	11	US-09-814-604-26
15	7	1.4	20	11	US-09-922-226-12

16	7	1.4	20	11	US-09-922-226-165	Sequence 165, App
17	7	1.4	28	15	US-09-982-172-226	Sequence 226, App
18	7	1.4	28	15	US-10-023-066A-58	Sequence 58, App
19	7	1.4	28	15	US-10-023-066A-73	Sequence 73, App
20	7	1.4	33	9	US-09-281-717-34	Sequence 34, App
21	7	1.4	35	9	US-09-864-761-37156	Sequence 37156, A
22	7	1.4	37	10	US-09-982-172-94	Sequence 94, App
23	7	1.4	45	15	US-10-074-475-243	Sequence 243, App
24	7	1.4	50	11	US-09-974-879-225	Sequence 225, App
25	7	1.4	51	11	US-09-105-736-225	Sequence 225, App
26	7	1.4	54	15	US-10-227-629-9	Sequence 9, App
27	7	1.4	60	11	US-09-809-991-403	Sequence 403, App
28	7	1.4	60	12	US-09-882-171-403	Sequence 403, App
29	7	1.4	68	14	US-10-001-857-130	Sequence 130, App
30	7	1.4	77	15	US-10-023-066A-75	Sequence 75, App
31	7	1.4	78	9	US-09-864-761-37650	Sequence 37650, A
32	7	1.4	86	9	US-09-859-101-4	Sequence 4, App
33	7	1.4	86	10	US-09-987-108-2	Sequence 2, App
34	7	1.4	86	10	US-09-987-108-4	Sequence 4, App
35	7	1.4	86	10	US-09-987-108-5	Sequence 5, App
36	7	1.4	86	10	US-09-987-108-6	Sequence 6, App
37	7	1.4	87	12	US-10-334-038-12	Sequence 12, App
38	7	1.4	104	9	US-09-859-101-3	Sequence 3, App
39	7	1.4	104	15	US-10-136-807A-6	Sequence 6, App
40	7	1.4	106	15	US-10-106-698-6734	Sequence 6734, App
41	7	1.4	111	9	US-09-925-299-1143	Sequence 1143, App
42	7	1.4	111	11	US-09-925-299-1143	Sequence 1143, App
43	7	1.4	158	12	US-10-032-585-7497	Sequence 7497, App
44	7	1.4	212	9	US-09-864-761-46554	Sequence 46554, A
45	7	1.4	275	9	US-09-815-242-10184	Sequence 10184, A

ALIGNMENTS

```
RESULT 1
US-09-938-803-6
; Sequence 6, Application US/09938803
; Patent No. US20020076762A1
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Tang, Y. Tom
; APPLICANT: Lal, Preeti
; APPLICANT: Reddy, Roopa
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Yang, Junning
; APPLICANT: Azimzal, Yalda
; TITLE OF INVENTION: FULL-LENGTH EXPRESSED GENETIC MARKERS
; FILE REFERENCE: PF-0695 US
; CURRENT APPLICATION NUMBER: US/09/938,803
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/311,894
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 6
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Inocyte Clone 1928920
US-09-938-803-6

Query Match      100.0%; Score 487, DB 9, Length 487,
Best Local Similarity 100.0%; Pred. No. 0,
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MASSAEDEGTVALAGVLSGFOELSLNKATSLGASQARLRIISIFLGPALFYRH 60
DB 1 MASSAEDEGTVALAGVLSGFOELSLNKATSLGASQARLRIISIFLGPALFYRH 60
QY YLFYKETYLIHLFHTFTGSLIAYFNFGNQLYHSLCLIVLPILRLMGRITTAVALTTFCF 120
```

Db 61 YIFVETVLIHLPHTFTGSIAYFNFQNLHSLCLIVQFILMLMGRTTIAVLTPFR 120
Qy 121 QMAVLLAGYATATGNYDIKMTMPCVLTLLKILGLAVDPDGKQNSLSSQKXAIRG 180
Db 121 QMAVLLAGYATATGNYDIKMTMPCVLTLLKILGLAVDPDGKQNSLSSQKXAIRG 180
Qy 181 VPSLLEVGAFSYFCAFLVGPQSMNHYKLVQGEILIDIPKIPNSIIPALRSLGLFY 240
Db 181 VPSLLEVGAFSYFCAFLVGPQSMNHYKLVQGEILIDIPKIPNSIIPALRSLGLFY 240
Qy 241 LVGYTLSPHITEDYLTEDYDNHPFRCMMLIMGKFLYKYVTGMLVTEGVCILTGL 300
Db 241 LVGYTLSPHITEDYLTEDYDNHPFRCMMLIMGKFLYKYVTGMLVTEGVCILTGL 300
Qy 301 GENGEEKGAACAKMDCANMKVLPETNPRTGTIASFNINAMAVARYFKRLKLGKKE 360
Db 301 GENGEEKGAACAKMDCANMKVLPETNPRTGTIASFNINAMAVARYFKRLKLGKKE 360
Qy 361 LSQGLSLFLALMHGHSGLVCPQMEFLIVVERQAALIQESPTLSKLAITVLOPRY 420
Db 361 LSQGLSLFLALMHGHSGLVCPQMEFLIVVERQAALIQESPTLSKLAITVLOPRY 420
Qy 421 YLVQOTIHLFMGYGNTAFCLFTWDMKLVKYSIYFLGHIFFLSLFIIPYIHKMVPK 480
Db 421 YLVQOTIHLFMGYGNTAFCLFTWDMKLVKYSIYFLGHIFFLSLFIIPYIHKMVPK 480
Qy 481 EKLKXME 487
Db 481 EKLKXME 487

RESULT 2

US-10-029-386-29742
; Sequence 29742, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEWICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 29742
; LENGTH: 57
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR12.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.5
; OTHER INFORMATION: SWISSPROT HIT: P54336, EVALU 8.00e+00
US-10-029-386-29742

Query Match 11.5%; Score 56; DB 12; Length 57;
Best Local Similarity 100.0%; Pred. No. 2.8e-46;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 292 EGVYCLITGIGFNGFEKGAACAKMDCANMKVLPETNPRTGTIASFNINAMAVAR 347
Db 2 EGVYCLITGIGFNGFEKGAACAKMDCANMKVLPETNPRTGTIASFNINAMAVAR 347

RESULT 3

US-09-864-761-33415
; Sequence 33415, Application US/09864761
; Patent No. US2002048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEWICA-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33415
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006512.12
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
; OTHER INFORMATION: EST_HUMAN HIT: W26367.1, EVALU 3.00e-26
; OTHER INFORMATION: SWISSPROT HIT: P43288, EVALU 4.50e+00
US-09-864-761-33415

Query Match 8.6%; Score 42; DB 9; Length 58;

Best Local Similarity 100.0%; Pred. No. 9,7e-33;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 QARLQESPTLSKLAITVLOPEFYLVQQTTHLPMGYSMY 437
Db 17 QARLQESPTLSKLAITVLOPEFYLVQQTTHLPMGYSMY 58

RESULT 4

US-09-859-101-5
; Sequence 5, Application US/09859101
; Patent No. US2002068829A1
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; Hillman, Jennifer L.
; Goli, Surya K.

TITLE OF INVENTION: NOVEL HUMAN DBI/ACBP-LIKE PROTEIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/859,101
FILING DATE: 14-May-2001

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/700,626
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0115 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-845-0555
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 86 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 118275

SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-859-101-5.

Query Match 1.6%; Score 8; DB 9; Length 86;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 KGKAKMDA 315
Db 50 KGKAKMDA 57

RESULT 5

US-09-987-108-1
; Sequence 1, Application US/09987108
; Patent No. US20020142347A1
; GENERAL INFORMATION:
; APPLICANT: KNUDSEN, Jens
; WADUM, Maiken C.T.
; APPLICANT: VILLADSEN, Jens

TITLE OF INVENTION: NOVEL HUMAN DBI/ACBP-LIKE PROTEIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: U.S.
ZIP: 94304

APPLICANT: NEERGAARD, Thomas B.F.
TITLE OF INVENTION: BIOSENSOR
FILE REFERENCE: KNUDSENIA

CURRENT APPLICATION NUMBER: US/09/987,108
CURRENT FILING DATE: 2001-11-13

PRIOR APPLICATION NUMBER: US 60/262,366
PRIOR FILING DATE: 2001-01-19

PRIOR APPLICATION NUMBER: DK PA2000 01638
PRIOR FILING DATE: 2000-11-10

NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn version 3.1

SEQ ID NO 1
LENGTH: 86
TYPE: PRT

ORGANISM: Bos taurus
US-09-987-108-1

Query Match 1.6%; Score 8; DB 10; Length 86;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 KGKAKMDA 315
Db 50 KGKAKMDA 57

RESULT 6

US-09-987-108-3
; Sequence 3, Application US/09987108
; Patent No. US20020142347A1
; GENERAL INFORMATION:
; APPLICANT: KNUDSEN, Jens
; WADUM, Maiken C.T.
; APPLICANT: VILLADSEN, Jens
; APPLICANT: NEERGAARD, Thomas B.F.

TITLE OF INVENTION: BIOSENSOR
FILE REFERENCE: KNUDSENIA
CURRENT APPLICATION NUMBER: US/09/987,108
CURRENT FILING DATE: 2001-11-13

PRIOR APPLICATION NUMBER: US 60/262,366
PRIOR FILING DATE: 2001-01-19

PRIOR APPLICATION NUMBER: DK PA2000 01638
PRIOR FILING DATE: 2000-11-10

NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn version 3.1

SEQ ID NO 3
LENGTH: 86
TYPE: PRT

ORGANISM: Sus scrofa
US-09-987-108-3

Query Match 1.6%; Score 8; DB 10; Length 86;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 KGKAKMDA 315
Db 50 KGKAKMDA 57

RESULT 7

US-09-987-108-7
; Sequence 7, Application US/09987108
; Patent No. US20020142347A1
; GENERAL INFORMATION:
; APPLICANT: KNUDSEN, Jens
; WADUM, Maiken C.T.
; APPLICANT: VILLADSEN, Jens
; APPLICANT: NEERGAARD, Thomas B.F.

TITLE OF INVENTION: BIOSENSOR
FILE REFERENCE: KNUDSENIA
CURRENT APPLICATION NUMBER: US/09/987,108
CURRENT FILING DATE: 2001-11-13

NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7
LENGTH: 86
TYPE: PRT
ORGANISM: Sus scrofa
US-09-987-108-7

;; PRIOR APPLICATION NUMBER: US 60/262,366
;; PRIOR FILING DATE: 2001-01-19
;; PRIOR APPLICATION NUMBER: DK PA2000 01638
;; PRIOR FILING DATE: 2000-11-10
;; NUMBER OF SEQ ID NOS: 38
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 7
;; LENGTH: 86
;; TYPE: PRT
;; ORGANISM: Terrestrial carolina
US-09-987-108-7

Query Match 1.6%; Score 8; DB 10; Length 86;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 KGRKAKWDA 315
DB 50 KGRKAKWDA 57

RESULT 8
US-09-987-108-8
; Sequence 8, Application US/09987108
; Patent No. US20020142347A1
; GENERAL INFORMATION:
; APPLICANT: KNUDSEN, Jens
; APPLICANT: WADUM, Maiken C.T.
; APPLICANT: VILLAGSEN, Jens
; APPLICANT: NEERGAARD, Thomas B.F.
; TITLE OF INVENTION: BIOSENSOR
; FILE REFERENCE: KNUDSENIA
; CURRENT APPLICATION NUMBER: US/09/987,108
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/262,366
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: DK PA2000 01638
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Anas platyrhynchos
US-09-987-108-8

Query Match 1.6%; Score 8; DB 10; Length 86;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 KGRKAKWDA 315
DB 50 KGRKAKWDA 57

RESULT 9
US-09-987-108-9
; Sequence 9, Application US/09987108
; Patent No. US20020142347A1
; GENERAL INFORMATION:
; APPLICANT: KNUDSEN, Jens
; APPLICANT: WADUM, Maiken C.T.
; APPLICANT: VILLAGSEN, Jens
; APPLICANT: NEERGAARD, Thomas B.F.
; TITLE OF INVENTION: BIOSENSOR
; FILE REFERENCE: KNUDSENIA
; CURRENT APPLICATION NUMBER: US/09/987,108
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/262,366
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: DK PA2000 01638
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 38

;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 9
;; LENGTH: 86
;; TYPE: PRT
;; ORGANISM: Gallus gallus
US-09-987-108-9

Query Match 1.6%; Score 8; DB 10; Length 86;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 KGRKAKWDA 315
DB 50 KGRKAKWDA 57

RESULT 10
US-09-987-108-20
; Sequence 20, Application US/09987108
; Patent No. US20020142347A1
; GENERAL INFORMATION:
; APPLICANT: KNUDSEN, Jens
; APPLICANT: WADUM, Maiken C.T.
; APPLICANT: VILLAGSEN, Jens
; APPLICANT: NEERGAARD, Thomas B.F.
; TITLE OF INVENTION: BIOSENSOR
; FILE REFERENCE: KNUDSENIA
; CURRENT APPLICATION NUMBER: US/09/987,108
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/262,366
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: DK PA2000 01638
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Rana sp.
US-09-987-108-20

Query Match 1.6%; Score 8; DB 10; Length 86;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 KGRKAKWDA 315
DB 50 KGRKAKWDA 57

RESULT 11
US-09-988-462-25
; Sequence 25, Application US/09988462
; Publication No. US20030046726A1
; GENERAL INFORMATION:
; APPLICANT: Kozziel, Michael G.
; Deesai, Nalin M.
; Lewis, Kelly S.
; Kramer, Vance C.
; Warren, Gregory W.
; Ewola, Stephen V.
; Crossland, Lyle D.
; Wright, Martha S.
; Merlin, Ellis J.
; Launus, Karen L.
; TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
; INSECTICIDAL ACTIVITY IN MAIZE
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Syngenta Biotechnology, Inc.
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC

COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/988,462
 FILING DATE: 20-NOV-2003
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 09/547,422
 FILING DATE: 11-APR-2000
 APPLICATION NUMBER: US 08/459,504
 FILING DATE: 02-JUN-1995
 APPLICATION NUMBER: US 07/951,715
 FILING DATE: 25-SEP-1992
 APPLICATION NUMBER: US 07/772,027
 FILING DATE: 04-OCT-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: S-188051
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919)541-8587
 TELEFAX: (919)541-8689
 INFORMATION FOR SEQ ID NO: 25:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 463 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 FEATURE:
 NAME/KEY: Protein
 LOCATION: 1..463
 OTHER INFORMATION: /notes "protein sequence for soybean CDPK as shown in Figure 34."
 SEQUENCE DESCRIPTION: SEQ ID NO: 25:
 US-09-988-462-25
 Query Match 1.6%; Score 8; DB 11; Length 463;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 394 ERQARLI 401
 Db 111 ERQARLI 118
 RESULT 12
 US-10-032-585-7294
 Sequence 7294, Application US/10032585
 Publication No. US20030180953A1
 GENERAL INFORMATION:
 APPLICANT: Terry, Roemer D.
 APPLICANT: Bo, Jiansheng
 APPLICANT: Charles, Boone
 APPLICANT: Howard, Bussey
 TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
 FILE REFERENCE: 10182-005-999
 CURRENT APPLICATION NUMBER: US/10/032,585
 CURRENT FILING DATE: 2001-12-20
 NUMBER OF SEQ ID NOS: 8000
 SOFTWARE: Patent in version 3.1
 SEQ ID NO 7294
 LENGTH: 609
 TYPE: PRT
 ORGANISM: Candida albicans
 US-10-032-585-7294

Query Match 1.6%; Score 8; DB 12; Length 609;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 296 ILTGIGFN 303
 Db 208 ILTGIGFN 215
 RESULT 13
 US-10-101-464A-73
 Sequence 73, Application US/10101464A
 Publication No. US20030046728A1
 GENERAL INFORMATION:
 APPLICANT: Strabala, Timothy
 APPLICANT: Nieuwenhuizen, Nicolaas
 APPLICANT: Higgins, Colleen M.
 TITLE OF INVENTION: Compositions Isolated from Plant Cells
 FILE REFERENCE: 11000.1020C2
 CURRENT APPLICATION NUMBER: US/10/101,464A
 CURRENT FILING DATE: 2002-03-18
 PRIOR APPLICATION NUMBER: 09/704,302
 PRIOR FILING DATE: 2000-11-01
 PRIOR APPLICATION NUMBER: 09/228,986
 PRIOR FILING DATE: 1999-01-12
 PRIOR APPLICATION NUMBER: 60/162,866
 PRIOR FILING DATE: 1999-11-01
 PRIOR APPLICATION NUMBER: PCT/US00/00724
 PRIOR FILING DATE: 2000-01-11
 NUMBER OF SEQ ID NOS: 989
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 73
 LENGTH: 947
 TYPE: PRT
 ORGANISM: Pinus radiata
 US-10-101-464A-73
 Query Match 1.6%; Score 8; DB 15; Length 947;
 Best Local Similarity 100.0%; Pred. No. 95;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 462 FLSLFIL 469
 Db 10 FLSLFIL 17
 RESULT 14
 US-09-814-604-26
 Sequence 26, Application US/09814604
 Publication No. US2003003517A1
 GENERAL INFORMATION:
 APPLICANT: Klein, Elliott S.
 APPLICANT: Chandraratna Roshantha A.
 TITLE OF INVENTION: Methods of Detecting Dissociated Nuclear
 FILE REFERENCE: P-AR 4528
 CURRENT APPLICATION NUMBER: US/09/814,604
 CURRENT FILING DATE: 2001-03-22
 NUMBER OF SEQ ID NOS: 52
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 26
 LENGTH: 20
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-814-604-26
 Query Match 1.4%; Score 7; DB 11; Length 20;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 76 FTGLSIA 82
 Db 76 FTGLSIA 82

Db 8 FTGLSIA 14

```

RESULT 15
US-09-922-226-12
; Sequence 12, Application US/099222226
; Publication No. US20030077664A1
; GENERAL INFORMATION:
; APPLICANT: Zhao, Yi
; APPLICANT: Thacher, Scott M.
; APPLICANT: Xiao, Jia-Hao
; APPLICANT: Kusari, Jyotirmoy
; APPLICANT: Chandraratna, Roshanra A.
; TITLE OF INVENTION: Methods of Screening For Compounds That
; TITLE OF INVENTION: Modulate Hormone Receptor Activity
; FILE REFERENCE: P-AR 4681
; CURRENT APPLICATION NUMBER: US/09/922,226
; PRIOR FILING DATE: 2002-01-09
; PRIOR APPLICATION NUMBER: US 60/284,797
; PRIOR FILING DATE: 2001-04-18
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-226-12
    
```

```

Query Match 1.4%; Score 7; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 76 FTGLSIA 82
DB 8 FTGLSIA 14
    
```

Search completed: November 21, 2003, 14:29:34
Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2003, 02:08:36 ; Search time 8288 Seconds

(without alignments) 11219.544 Million cell updates/sec

Title: US-09-938-803-26

Perfect score: 2273
Sequence: 1 ggggggtgaagcagcagctt.....ttttcaaaaaaaaaaaaaa 2273

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GeneB1: *
1: gb_ba: *
2: gb_hgt: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pac: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_scs: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pac: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_scs: *
28: em_un: *
29: em_vl: *
30: em_hgt_hum: *
31: em_hgt_inv: *
32: em_hgt_other: *
33: em_hgt_mus: *
34: em_hgt_pin: *
35: em_hgt_rtd: *
36: em_hgt_mam: *
37: em_hgt_vrt: *
38: em_hgt_sy: *
39: em_hgt_hum: *
40: em_hgt_mus: *
41: em_hgt_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2273	100.0	2273	6 AX048067	AX048067 Sequence
2	2241.2	98.6	2264	6 AX073570	AX073570 Sequence
3	2215	97.4	2482	6 AK096775	AK096775 Homo sapi
4	2045.4	90.0	2058	6 BC000664	BC000664 Homo sapi
5	1842	81.0	1842	6 AX331926	AX331926 Sequence
6	1842	81.0	1842	6 AX410769	AX410769 Sequence
7	1842	81.0	1842	6 HSU72515	U72515 Human C3f m
8	1461	64.3	1461	6 AX073560	AX073560 Sequence
9	1388.2	61.1	1573	6 AK058063	AK058063 Homo sapi
10	1240.4	54.6	2128	10 AK098090	AK098090 Mus muscu
11	1233.4	54.3	1898	10 BC006753	BC006753 Mus muscu
12	1229.4	54.1	1968	10 AY028317	AY028317 Mus muscu
13	1143.4	50.3	1146	9 BT007000	BT007000 Homo sapi
14	1142.4	50.3	1146	12 BT007735	BT007735 Synthetic
15	708.6	31.2	155975	9 AC006512	AC006512 Homo sapi
16	708.6	31.2	222930	9 HSU47924	U47924 Human chrom
17	433.2	19.1	450	6 BD023986	BD023986 Sequence
18	425.4	18.7	1641	6 BC009856	BC009856 Homo sapi
19	254.8	11.2	990	6 AX331531	AX331531 Sequence
20	254.8	11.2	990	6 AX331927	AX331927 Sequence
21	254.8	11.2	990	6 HSU72517	U72517 Human alter
22	249.6	11.0	227538	10 AC002397	AC002397 Mouse chr
23	249.6	11.0	227538	10 AY050668	AY050668 Homo sapi
24	236.4	10.4	169660	9 AC011479	AC011479 Homo sapi
25	236.4	10.4	169661	9 AC011479	AC011479 Homo sapi
26	235.6	10.4	192440	9 CENS01DXH	AL139316 Human chr
27	233.2	10.3	176541	9 AL329325	AL329325 Human DNA
28	233	10.3	140348	9 AL390725	AL390725 Human DNA
29	233	10.3	321519	2 AL714004	AL714004 Homo sapi
30	232.8	10.2	147054	2 AL356582	AL356582 Homo sapi
31	232.4	10.2	67885	2 AC107924	AC107924 Homo sapi
32	232.4	10.2	86156	2 AL162262	AL162262 Homo sapi
33	232.4	10.2	115093	2 AC091558	AC091558 Homo sapi
34	232.2	10.2	146350	9 HS22515	AL773604 Homo sapi
35	232.2	10.2	340000	9 HS21C100	AL163300 Homo sapi
36	232	10.2	321519	2 AL714004	AL714004 Homo sapi
37	231.2	10.2	167548	9 AL391809	AL391809 Human DNA
38	231	10.2	128829	9 AC127002	AC127002 Homo sapi
39	231	10.2	147419	9 HS738P1	AL031736 Human DNA
40	231	10.2	194296	9 AL354864	AL354864 Human DNA
41	231	10.2	201508	2 AC026290	AC026290 Homo sapi
42	230.8	10.2	37680	2 AL359985	AL359985 Homo sapi
43	230.8	10.2	93582	2 AC090681	AC090681 Homo sapi
44	230.8	10.2	146017	2 AC027473	AC027473 Homo sapi
45	230.8	10.2	151465	2 AC117371	AC117371 Homo sapi

ALIGNMENTS

RESULT 1
AX048067
LOCUS AX048067 2273 bp DNA linear PAT 15-DEC-2000
DEFINITION Sequence 61 from Patent WO0070047.
ACCESSION AX048067
VERSION AX048067.1 GI:11876890
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 Yue, H., Tang, Y. T., Lai, P., Reddy, R., Batra, S., Baughn, M. R.,
Yang, J., Azimzai, Y., Lu, D. A., Au-Young, J. and Shi, L. L.,
Full-length molecules expressed in human tissues

Pred. No. is the number of results predicted by chance to have a

JOURNAL Patent: WO 0070047-A 61.23-NOV-2000;
Incyle Genomics, Inc. (US)
Location/Qualifiers
FEATURES
source
1..2273
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="Incyle ID No: 1929920CB1"
BASE COUNT 494 a 630 c 547 g 602 t
ORIGIN
Query Match 100.0%; Score 2273; DB 6; Length 2273;
Best Local Similarity 100.0%; Pred. No.0;
Matches 2273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGGGTGAAGCGATACGTTTGGCCGCAATCGGGGGCGCGGAGCTGGGGGGTCCCTGTG 60
Db 1 GGGGGTGAAGCGATACGTTTGGCCGCAATCGGGGGCGCGGAGCTGGGGGGTCCCTGTG 60
QY 61 GGGCTCCCGAGTTAAGATGGCGTCTCAGCGAGGGGGAGAGAGGGAGCTGTGGTGGCG 120
Db 61 GGGCTCCCGAGTTAAGATGGCGTCTCAGCGAGGGGGAGAGAGGGAGCTGTGGTGGCG 120
QY 121 TGGCGGGGGTTCGACGTGGGTTTCCAGGAGCTGAAGCTTAAACAAGTTGGCGAGCTCC 180
Db 121 TGGCGGGGGTTCGACGTGGGTTTCCAGGAGCTGAAGCTTAAACAAGTTGGCGAGCTCC 180
QY 181 TGGCGCGTCAGAAACAGGCGCTCGGCTGATCATCTTCATCTTCTGGGTTACCCCTTG 240
Db 181 TGGCGCGTCAGAAACAGGCGCTCGGCTGATCATCTTCATCTTCTGGGTTACCCCTTG 240
QY 241 CTTTGTGTTTATCGGCATTAACCTTTTCTACAAGAGACTTACCTCATCCTCTTCCATA 300
Db 241 CTTTGTGTTTATCGGCATTAACCTTTTCTACAAGAGACTTACCTCATCCTCTTCCATA 300
QY 301 CCTTACAGGCGCTCTCAATGCTTATTTAATTTGGAACAGAGCTTACAGCTCCCTGC 360
Db 301 CCTTACAGGCGCTCTCAATGCTTATTTAATTTGGAACAGAGCTTACAGCTCCCTGC 360
QY 361 TGTGATGTCCTTCAAGTTCCTCATCTTGAATAATGGGCGGACCATCATCGCGCTC 420
Db 361 TGTGATGTCCTTCAAGTTCCTCATCTTGAATAATGGGCGGACCATCATCGCGCTC 420
QY 421 TCACTACCTTTTGGCTTCCAGATGGCTTCTGGCTGATTAATTAATTCAGCCAGC 480
Db 421 TCACTACCTTTTGGCTTCCAGATGGCTTCTGGCTGATTAATTAATTCAGCCAGC 480
QY 481 GCAACTACGATATCAAGTGAAGTGCACATGTTGTTCTGACTTTGAAGCTGATTGGT 540
Db 481 GCAACTACGATATCAAGTGAAGTGCACATGTTGTTCTGACTTTGAAGCTGATTGGT 540
QY 541 TGGCTGTGATCTTGTGACGAGGAGGAAAGATCAGAAATTCCTGTCTCTGAGCAACAG 600
Db 541 TGGCTGTGATCTTGTGACGAGGAGGAAAGATCAGAAATTCCTGTCTCTGAGCAACAG 600
QY 601 AATATCCATACCTGCTGTTCTCTTCCCTGCTGAAGTGTGTTCTCTTACTTTCTAG 660
Db 601 AATATCCATACCTGCTGTTCTCTTCCCTGCTGAAGTGTGTTCTCTTACTTTCTAG 660
QY 661 GGGCTTCTTGTAGGGGCCCAAGTTCATGATGATCACTACATGAAGCTGTGAGGAG 720
Db 661 GGGCTTCTTGTAGGGGCCCAAGTTCATGATGATCACTACATGAAGCTGTGAGGAG 720
QY 721 AGCTGATGACATACGAGAAAGATACCAACAGCATCATTCCTGCTCTCAAGCGCTGA 780
Db 721 AGCTGATGACATACGAGAAAGATACCAACAGCATCATTCCTGCTCTCAAGCGCTGA 780
QY 781 GTCTGGGCTTTTCTACTAGTGGGTACACACTGCTCAGCCCCACATACAGAAAGACT 840
Db 781 GTCTGGGCTTTTCTACTAGTGGGTACACACTGCTCAGCCCCACATACAGAAAGACT 840
QY 841 ATCTCCCTACTGAAGACTATGACCAACCCCTTCTGGTTCGGGTGATTAATGCTGA 900
Db 841 ATCTCCCTACTGAAGACTATGACCAACCCCTTCTGGTTCGGGTGATTAATGCTGA 900

Db 841 ATCTCCCTACTGAAGACTATGACCAACCCCTTCTGGTTCGGGTGATTAATGCTGA 900
QY 901 TCTGGGGCAAGTTTGTGCTGTACAATATATGCACTGTTGGCTGTACAGAGAGATAT 960
Db 901 TCTGGGGCAAGTTTGTGCTGTACAATATATGCACTGTTGGCTGTACAGAGAGATAT 960
QY 961 GCATTTTGAAGGGCCCTGGCTTCAATGCTTTGAAGAAAAGGCGAAGGCTGGGATG 1020
Db 961 GCATTTTGAAGGGCCCTGGCTTCAATGCTTTGAAGAAAAGGCGAAGGCTGGGATG 1020
QY 1021 CCTGTGCAACATGAAGTGTGGCTCTTGAAGAAAAGGCGCTTCACTGAGCAACATTG 1080
Db 1021 CCTGTGCAACATGAAGTGTGGCTCTTGAAGAAAAGGCGCTTCACTGAGCAACATTG 1080
QY 1081 CCTCATTTCAACATCAACACCAAGCGCTGGGTGGCCGCTACATCTTCAACAGCTCAAGT 1140
Db 1081 CCTCATTTCAACATCAACACCAAGCGCTGGGTGGCCGCTACATCTTCAACAGCTCAAGT 1140
QY 1141 TCTTGGAAATTAAGAACTCTCAGAGTCTCTGTTGCTATTTCTGGCCCTCTGGCAG 1200
Db 1141 TCTTGGAAATTAAGAACTCTCAGAGTCTCTGTTGCTATTTCTGGCCCTCTGGCAG 1200
QY 1201 GCCTGACTCAGGATACCTGTGCTCTGCTTCCAGATGGAATTCCTCATTTGTTATGGAAA 1260
Db 1201 GCCTGACTCAGGATACCTGTGCTCTGCTTCCAGATGGAATTCCTCATTTGTTATGGAAA 1260
QY 1261 GACAGGCTCCAGGCTCATTTCAAGAGGCCCACTGTAGCAAGCTGGCCGCTATTACTG 1320
Db 1261 GACAGGCTCCAGGCTCATTTCAAGAGGCCCACTGTAGCAAGCTGGCCGCTATTACTG 1320
QY 1321 TCTCCAGGCTTCTTCTATTTTGGTGAACAACAACATCACTGGCTTCAAGGGTTACT 1380
Db 1321 TCTCCAGGCTTCTTCTATTTTGGTGAACAACAACATCACTGGCTTCAAGGGTTACT 1380
QY 1381 CCATGACTGCTTCTGCTCTTCACTGAGGAAATGCTTAAGGTGATTAATCCATCT 1440
Db 1381 CCATGACTGCTTCTGCTCTTCACTGAGGAAATGCTTAAGGTGATTAATCCATCT 1440
QY 1441 ATTTCTTGGCAATCTTCTCTGAGCTTCAATTAATGCTTATTAATTCACAAAG 1500
Db 1441 ATTTCTTGGCAATCTTCTCTGAGCTTCAATTAATGCTTATTAATTCACAAAG 1500
QY 1501 CAATGTCGCAAGAAAGAAAGTAAAGAAATGAATGAATTCATTCCTCTGGTGGCT 1560
Db 1501 CAATGTCGCAAGAAAGAAAGTAAAGAAATGAATGAATTCATTCCTCTGGTGGCT 1560
QY 1561 GTGCGGAACTGTGTCAGAAACTACTGCTCTCTTTCACAGCACTCTTGGCCGAGAG 1620
Db 1561 GTGCGGAACTGTGTCAGAAACTACTGCTCTCTTTCACAGCACTCTTGGCCGAGAG 1620
QY 1621 CAGAGATGAAAAGCAGAGGAGTGAAGATGAGTCTTCCAGCTGTGCTCTGCTGCG 1680
Db 1621 CAGAGATGAAAAGCAGAGGAGTGAAGATGAGTCTTCCAGCTGTGCTCTGCTGCG 1680
QY 1681 AGCAAGTCTTCAATTTGGGGCCAAAAGGGGAACTTTTGTGAGAAAGGCGCTTGGCTT 1740
Db 1681 AGCAAGTCTTCAATTTGGGGCCAAAAGGGGAACTTTTGTGAGAAAGGCGCTTGGCTT 1740
QY 1741 GTTACCCACGCTGATGAGTGGCGGATCTCAGCTCAACCGCACTCTCCTCTGGG 1800
Db 1741 GTTACCCACGCTGATGAGTGGCGGATCTCAGCTCAACCGCACTCTCCTCTGGG 1800
QY 1801 TTCAAGTGAATTTTCCGCTCAGCTCCCAAGTGTGGGAAATACAGGACGCAACATG 1860
Db 1801 TTCAAGTGAATTTTCCGCTCAGCTCCCAAGTGTGGGAAATACAGGACGCAACATG 1860
QY 1861 CCAGCTAATTTTGTGATTTTCAAGTGAAGAGGGAATTCACAGTGTGGCGAGGCTGTC 1920
Db 1861 CCAGCTAATTTTGTGATTTTCAAGTGAAGAGGGAATTCACAGTGTGGCGAGGCTGTC 1920
QY 1921 TCGAATCTCTGACCGCAAGTGAATCAACCGGCTCTGCGCTCCCAAGTGTGGGATTAACAG 1980
Db 1921 TCGAATCTCTGACCGCAAGTGAATCAACCGGCTCTGCGCTCCCAAGTGTGGGATTAACAG 1980

QY	1991	GCGTGAAGCACCGTCCCGCGCCGAAGAAGGAAACCTCTGTGTGGAGAGACACAGAGGAGCTCA	2040
Db	1991	GCGTGAAGCACCGTCCCGCGCCGAAGAAGGAAACCTCTGTGTGGAGAGACAGAGGAGGCTCA	2040
QY	2041	CATCTCCCGCTGTATTCCGCCATGCACATTCGCTATCTCCGCATTCAGCCAGATC	2100
Db	2041	CATCTCCCGCTGTATTCCGCCATGCACATTCGCTATCTCCGCATTCAGCCAGATC	2100
QY	2101	TATTGTGTTTTCTTCTTCGCCAATTACATATGATTGTGATGTGCGCGCTACCAACCAACCCCC	2160
Db	2101	TATTGTGTTTTCTTCTTCGCCAATTACATATGATTGTGATGTGCGCGCTACCAACCAACCCCC	2160
QY	2161	CCCATGGGGGGGTGAGAGAGGGGTGCAGAGGCCCTGCTCCTCACTTTTCTACCTTGGAA	2220
Db	2161	CCCATGGGGGGGTGAGAGAGGGGTGCAGAGGCCCTGCTCCTCACTTTTCTACCTTGGAA	2220
QY	2221	CTGTATTGATTAATACTCTCTGTGTGTTCAGTTTTCACAAAAA	2273
Db	2221	CTGTATTGATTAATACTCTCTGTGTGTTCAGTTTTCACAAAAA	2273

RESULT 2					
AX073570	AX073570	2264 bp	DNA	linear	PAT 06-FEB-2001
LOCUS	Sequence 22	from Patent WO0104297.			
DEFINITION	AX073570				
ACCESSION	AX073570.1	GI:12709984			
VERSION					

SOURCE ORGANISM	Homo sapiens (human)
ORGANISM	Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

· AUTHORS
Kato, S. and Kimura, T.
TITLE
Human protein having hydrosphobic domain and drug conjugating character

JOURNAL Patent: WO 0104297-A 22 18-JAN-2001;
SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Protegene Inc. (JP)
FEATURES Location/Qualifiers

```
source      1. .2264
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
CDS        85. .1548
```

```

/note="unnamed protein product"
/codon_start=1
/protein_id="CAC28402.1"
/db_xref="GI:12709985"
/translation="MASSAGEDETECTVVALAGVLOSQFQELSLNKLATSLGASGQALH
IISFLGPPALFYRHFALFYKFTYLLIHLFTFTGSLSTATPNRGNQLYHSLCTVLG
ILRMGRITIAVLITTFPCQMAVYLAAGYVYATNDRIKTMPEHCVLTLLKLGAVLV
DGGQDNQSSLSQOQYAIRGVPSYLSVAGSYSGAFVGPQSNMHWKLVLOGEIE
IPGKIPNSIIPALKRILSLGFLVGYTLITLEDYDNDHPFWRFCMYMVI
GKFLYKVTQCMILVTBEGVCLITGLGFGFGEPEKKAQWMDCAANKWLFPTNRFPG
ASEFINNTAWAVARIERKLKFLGNKELSQSLSLFLALMGHLSGVLVCFQMFLL
VERQARLIGSSPTLSKLAIVLQPIYVLOQTIMLFWGYSMTAFCLFTWDKWL
YKSYTFIGHIFSLPLTPIYIHKAMPREKRLKQME"

```

[illegible]

QY	120	TTGGCGGGGGGTTCTGCAGTCGGGTTTCCAGAGACTGAGCCCTTAACAAGTTGGCGAGCTCC	179
Db	127	CTGGGGGGGGGTTCTGCGACAGTCGGGTTTTCCAGGAGCTGAGCCTTTAACAGATTGGCGACGGTCC	186
QY	180	CTGGGCGGGTCAAGAACAGGCGCTCGGGGTATCATCTTCACCTTCTCGGGTTACCCCTTT	239
Db	187	CTGGGCGGGTCAAGAACAGGCGCTCGGGGTATCATCTTCACCTTCTCGGGTTACCCCTTT	246
QY	240	GCTTTGTTTTATCGGCAATTACCTTTTCTACAAGAGACTACCTATCCACCTCTTCCAT	299
Db	247	GCTTTGTTTTATCGGCAATTACCTTTTCTACAAGAGACTACCTATCCACCTCTTCCAT	306
QY	300	ACCTTTACAGGCTCTCAATTGCTTATTTTACTTTGAAAACAGGCTTACCACCTCCGTG	359
Db	307	ACCTTTACAGGCTCTCAATTGCTTATTTTAACTTTGAAAACAGGCTTACCACCTCCGTG	366
QY	360	CTGGTATATGCTTCAGTTCCTCATCCTTCAGCTAATGGGCGGACCATCAGTCGGCTC	419
Db	367	CTGGTATATGCTTCAGTTCCTCATCCTTCAGCTAATGGGCGGACCATCAGTCGGCTC	426
QY	420	CTCACTACCTTTTGCTTCCAGATAGGCGCTTCTGCTGATGATCTATTAACACTGCCACC	479
Db	427	CTCACTACCTTTTGCTTCCAGATAGGCGCTTCTGCTGATGATCTATTAACACTGCCACC	486
QY	480	GCGCACTACGATATCAGTGAAGTGCACATGCTGTGTTCTGACTTTGAGCTGATTTGGT	539
Db	487	GCGCACTACGATATCAGTGAAGTGCACATGCTGTGTTCTGACTTTGAGCTGATTTGGT	546
QY	540	TTGGCTGTGCTACTTTTGAAGGAGGGGAAATAGATTCCTTGTCCTCGAGCAACAG	599
Db	547	TTGGCTGTGCTACTTTTGAAGGAGGGGAAATAGATTCCTTGTCCTCGAGCAACAG	606
QY	600	AAATATGCGCAACAGCTGATGTTCTCTTCCCTGCTGGAAGTTGCTGGTTTCTCCTACTTCTAT	659
Db	607	AAATATGCGCAACAGCTGATGTTCTCTTCCCTGCTGGAAGTTGCTGGTTTCTCCTACTTCTAT	666
QY	660	GGGGCTTCTTGGTAGGGCCCCAGTTCTCAATGAATCACTACATGAAGCTGTGCAGGGA	719
Db	667	GGGGCTTCTTGGTAGGGCCCCAGTTCTCAATGAATCACTACATGAAGCTGTGCAGGGA	726
QY	720	GAGCTGATGTGCATACCAAGGAAATATCCAAACAGCATCTCCTGCTCTCAAGCGCTG	779
Db	727	GAGCTGATGTGCATACCAAGGAAATATCCAAACAGCATCTCCTGCTCTCAAGCGCTG	786
QY	780	AGTCTGGGCTCTTTCTACTAGTGGGCTACACACTGCTCAGCCCCCACAATCAGAGAGAC	839
Db	787	AGTCTGGGCTCTTTCTACTAGTGGGCTACACACTGCTCAGCCCCCACAATCAGAGAGAC	846
QY	840	TATCTCTCACTGAGAGCTATGACACCAACCCTTCTGGTTCGCGTGCATGTACATGCTG	899
Db	847	TATCTCTCACTGAGAGCTATGACACCAACCCTTCTGGTTCGCGTGCATGTACATGCTG	906
QY	900	ATCTGGGGCAAGTTTGCTGTATCAATAATGTCACTGTGGCTGTCAAGAGAGATA	959
Db	907	ATCTGGGGCAAGTTTGCTGTATCAATAATGTCACTGTGGCTGTCAAGAGAGATA	966
QY	960	TGCATTTTGAAGGCGCTGAGCTTCAATGGCTTTGAAGAAAAGGCGAAGGCAAGTGGAT	1019
Db	967	TGCATTTTGAAGGCGCTGAGCTTCAATGGCTTTGAAGAAAAGGCGAAGGCAAGTGGAT	1026
QY	1020	GCTGTGTGCCAACATGAAGTGTGGCTTTTGAAAACAAACCCCGCTTCACTGGGACATTT	1079
Db	1027	GCTGTGTGCCAACATGAAGTGTGGCTTTTGAAAACAAACCCCGCTTCACTGGGACATTT	1086
QY	1080	GCTTCATTCAACATCAACACCAAGGCTGTGGTGGCCGCTACATCTTCAACAGACTCAAG	1139
Db	1087	GCTTCATTCAACATCAACACCAAGGCTGTGGTGGCCGCTACATCTTCAACAGACTCAAG	1146
QY	1140	TTCTCTGGAAATTAAGAACTCTCTCAGAGGTCTCTGTTGCTATTTCTGGGCGCTCTGGAC	1199
Db	1147	TTCTCTGGAAATTAAGAACTCTCTCAGAGGTCTCTGTTGCTATTTCTGGGCGCTCTGGAC	1206
QY	1200	GCGCTGCACTCAGATACCTGGTCTGGTCTCCAGATGGAATTCCTCATTTGTTATGTGGA	1259

Db 1207 GGGCTGACCTGAGAACTGCTGCTTCCAGATGAAATCTCATGTATTTGGAA 1266
 Qy 1260 AGACAGGCTGCCAGGCTCATTTCAAGAGAGCCCACTGAGCAAGCTGGCCGCTTACT 1319
 Db 1267 AGACAGGCTGCCAGGCTCATTTCAAGAGAGCCCACTGAGCAAGCTGGCCGCTTACT 1326
 Qy 1320 GTTCTCCAGCCCTTCTACTATTTGGTGCACAGACATCATCTGCTCTTCAAGGTTAC 1379
 Db 1327 GTTCTCCAGCCCTTCTACTATTTGGTGCACAGACATCATCTGCTCTTCAAGGTTAC 1386
 Qy 1380 TCCATACGCTGCTTCTGCTCTTCAAGTGCACAAATGCTTAAAGTAAATCATC 1439
 Db 1387 TCCATACGCTGCTTCTGCTCTTCAAGTGCACAAATGCTTAAAGTAAATCATC 1446
 Qy 1440 TATTTCTTGGCCACATCTTCTTCTGAGCTCATTTCAATATGCTTATTTCAAA 1499
 Db 1447 TATTTCTTGGCCACATCTTCTTCTGAGCTCATTTCAATATGCTTATTTCAAA 1506
 Qy 1500 GCATATGTCGCAAGAAAGAAAGTAAAGAGATGATATTCATTTCTGCTGGCC 1559
 Db 1507 GCATATGTCGCAAGAAAGAAAGTAAAGAGATGATATTCATTTCTGCTGGCC 1566
 Qy 1560 TGTGCGGAGCTGTGAGAAACTACTGCTCTTCTTCAAGCACTCTTGGCCGAG 1619
 Db 1567 TGTGCGGAGCTGTGAGAAACTACTGCTCTTCTTCAAGCACTCTTGGCCGAG 1626
 Qy 1620 GCAGAGATGAAAGCAAGGAGAGTGAAGATGATGCTTCCAGCTGTGCTCTGCTG 1679
 Db 1627 GCAGAGATGAAAGCAAGGAGAGTGAAGATGATGCTTCCAGCTGTGCTCTGCTG 1686
 Qy 1680 CAGCCAACTTCTATTTGGGGCCAAAGGGAACTTTTGGAGAGGGCTTGTCTT 1739
 Db 1687 CAGCCAACTTCTATTTGGGGCCAAAGGGAACTTTTGGAGAGGGCTTGTCTT 1746
 Qy 1740 TGTACCCAGCTGAGATGAGTGGCGGATCTCAGCTCACCGCACTCTGCTGCTG 1799
 Db 1747 TGTACCCAGCTGAGATGAGTGGCGGATCTCAGCTCACCGCACTCTGCTGCTG 1806
 Qy 1800 GTTCAAGTATTTCTGCTGAGCTCTCCCAAGTACTGAGAAATACAGGAGCCCACT 1859
 Db 1807 GTTCAAGTATTTCTGCTGAGCTCTCCCAAGTACTGAGAAATACAGGAGCCCACT 1866
 Qy 1860 GCGGAGTATTTTGTATTTTCAATGAAACGGGATTTTCAACAGTGTGCTGCTGCT 1919
 Db 1867 GCGGAGTATTTTGTATTTTCAATGAAACGGGATTTTCAACAGTGTGCTGCTGCT 1926
 Qy 1920 CTGCAACTCTGACCGGAGTATTCACCGGCTCTGCTCTCCCAAGTGTGAGGATTA 1979
 Db 1927 CTGCAACTCTGACCGGAGTATTCACCGGCTCTGCTCTCCCAAGTGTGAGGATTA 1986
 Qy 1980 GGGGTGAGCCAGCTGCGCCGCCCAAGGGGAACTTTTGGGAGAGAGAGGGGCTC 2039
 Db 1987 GGGGTGAGCCAGCTGCGCCGCCCAAGGGGAACTTTTGGGAGAGAGAGGGGCTC 2046
 Qy 2040 ACATCTTCTCTGATTTCCCAATGCAATGCTTATCTCTCCCAATCTAGCCAGGAAT 2099
 Db 2047 ACATCTTCTCTGATTTCCCAATGCAATGCTTATCTCTCCCAATCTAGCCAGGAAT 2106
 Qy 2100 CTATGTGTTTCTTCTGCAATTTTACTATGATGATGATGATGATGATGATGATGAT 2159
 Db 2107 CTATGTGTTTCTTCTGCAATTTTACTATGATGATGATGATGATGATGATGATGAT 2166
 Qy 2160 CCGCATGAGGGGGTGAAGAGGGGTGAAGGCGCTGCTGCTCACTTTTCTACTTTGA 2219
 Db 2167 CCGCATGAGGGGGTGAAGAGGGGTGAAGGCGCTGCTGCTCACTTTTCTACTTTGA 2226
 Qy 2220 ACTGATATTAATAATCACTTCTGTTTGTTCAGTTT 2257
 Db 2227 ACTGATATTAATAATCACTTCTGTTTGTTCAGTTT 2264

RESULT 3

AK096775 2482 bp mRNA linear PRI 15-JUL-2002
 LOCUS AK096775 Homo sapiens cDNA FLJ39456 fis, clone PROST2010782, highly similar
 DEFINITION to Human C3f mRNA.
 ACCESSION AK096775
 VERSION AK096775.1 GI:21756342
 KEYWORDS oligo cloning; file (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE
 AUTHORS 1
 Nimmiya, K., Magatsuma, M., Kanda, K., Kondo, H., Yokoi, T.,
 Kodaira, H., Furuya, T., Takahashi, M., Kikawa, E., Onura, Y., Abe, K.,
 Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M.,
 Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,
 Yamamoto, J., Isono, Y., Kawai-Hiro, Y., Saito, K., Nishikawa, T.,
 Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M.,
 Kikuchi, H., Murakawa, K., Kanehori, K., Takahashi-Fujii, A.,
 Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,
 Nagahara, K., Masuno, Y., Nagai, K., and Isogai, T.
 NEDO human cDNA sequencing project
 TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2482)
 Isogai, T., and Yamamoto, J.
 Direct Submission
 AUTHORS Isogai, T., and Yamamoto, J.
 TITLE Direct Submission
 JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
 (E-mail: genom.cse@ri.ri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan: cDNA full insert sequencing:
 Research Association for Biotechnology (RAB); cDNA library
 construction: Helix Research Institute (HRI) (supported by Japan
 Key Technology Center etc.); 5- & 3- end one pass sequencing: RAB,
 HRI, and Biotechnology Center, National Institute of Technology and
 Evaluation; clone selection for full insert sequencing: HRI and
 RAB; annotation: HRI and RAB.
 FEATURES
 source
 1..2482
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="PROST2010782"
 /tissue_type="prostate"
 /clone_id="PROST2"
 /note="cloning vector: pME18FLJ3"
 BASE COUNT 526 a 687 c 615 g 654 t
 ORIGIN
 Query Match 97.4%; Score 2215; DB 9; Length 2482;
 Best Local Similarity 99.6%; Pred. No. 0; Mismatches 5; Indels 3; Gaps 3;
 Matches 2251; Conservative 0;
 Qy 1 GGGGGTGAAGCATTA-CGTTTGGCCGCAATTCGGGGCGCGGAGTCTGGGGGGTCCCTGT 59
 Db 226 GGGGGTGAAGCATTAAGCTTTTCCCGCATTCGGGGCGCGGAGTCTGGGGGGTCCCTGT 285
 Qy 60 GGGGCTCCGGAATTAAGATGCGTCTTCAGCGGAGGGGAGCAAGGGGACTGTGTGGCG 119
 Db 286 GGGGCTCCGGAATTAAGATGCGTCTTCAGCGGAGGGGAGCAAGGGGACTGTGTGGCG 345
 Qy 120 CTGGCGGGGGTCTGAGATCGGGTTTCCAGAGCTGAGAGCTTAACAAGTGGGAGTGC 179
 Db 346 CTGGCGGGGGTCTGAGATCGGGTTTCCAGAGCTGAGAGCTTAACAAGTGGGAGTGC 405
 Qy 180 CTGGGCGCTCAAGACAGGCGCTGCGGCTGATCATCTTCATCTTCTGGGTTACCCCTTT 239
 Db 406 CTGGGCGCTCAAGACAGGCGCTGCGGCTGATCATCTTCATCTTCTGGGTTACCCCTTT 465
 Qy 240 GCTTTGTTTATGCGCATTAACCTTTTCTACAAAGAGACCTTACTTCACTCTTTCAT 299
 Db 466 GCTTTGTTTATGCGCATTAACCTTTTCTACAAAGAGACCTTACTTCACTCTTTCAT 525

QY 300 ACCTTTACAGGCGCTCTCAATGCTATTATTTAACTTTGAAACCAAGCTTACCACTCCCTG 359
 DB 526 ACCTTTACAGGCGCTCTCAATGCTATTATTTAACTTTGAAACCAAGCTTACCACTCCCTG 585
 QY 360 CTGTGATTGTGCTTCAAGTTCTCTCATCTTCCATGATAGGCGGCAACCATCACTGCGCTG 419
 DB 586 CTGTGATTGTGCTTCAAGTTCTCTCATCTTCCATGATAGGCGGCAACCATCACTGCGCTG 645
 QY 420 CTCACTACCTTTGCTTCAGATGGCTTCTTCCGCTGAGATGATGATTTACACTGCGAC 479
 DB 646 CTCACTACCTTTGCTTCAGATGGCTTCTTCCGCTGAGATGATGATTTACACTGCGAC 705
 QY 480 GGCACATAGATATCAAGTGAAGCAATGACCAATGCTGCTGATGATGATGATGATGATGAT 539
 DB 706 GGCACATAGATATCAAGTGAAGCAATGCTGCTGATGATGATGATGATGATGATGATGAT 765
 QY 540 TTGGCTGTGATCTACTTTTACGAGGAGGAAAGATCAGAAATTCCTGCTGCTGAGCAAC 599
 DB 766 TTGGCTGTGATCTACTTTTACGAGGAGGAAAGATCAGAAATTCCTGCTGCTGAGCAAC 825
 QY 600 AAATATGACATAGCTGCTGCTTCTTCCGCTGAGATGATGATGATGATGATGATGATGAT 659
 DB 826 AAATATGACATAGCTGCTGCTTCTTCCGCTGAGATGATGATGATGATGATGATGATGAT 885
 QY 660 GGGGCGCTTCTTGTAGGCGCCCAAGTTCTCAATGATCACTAGATGATGATGATGATGAT 719
 DB 886 GGGGCGCTTCTTGTAGGCGCCCAAGTTCTCAATGATCACTAGATGATGATGATGATGAT 945
 QY 720 GAGCTGATGATATACAGGAAAGATACCAACAGATATCTGCTGCTGAGGCGCTG 779
 DB 946 GAGCTGATGATATACAGGAAAGATACCAACAGATATCTGCTGCTGAGGCGCTG 1005
 QY 780 AGCTGCGGCTTTTCTACCTAGGCGCTACACTGCTGAGGCGCCCACTGAGCAAGAC 839
 DB 1006 AGCTGCGGCTTTTCTACCTAGGCGCTACACTGCTGAGGCGCCCACTGAGCAAGAC 1065
 QY 840 TATCTCTCTCACTGAGACTATGACCAACACCCCTTCTGCTGCTGCTGATGATGATGAT 899
 DB 1066 TATCTCTCTCACTGAGACTATGACCAACACCCCTTCTGCTGCTGCTGATGATGATGAT 1125
 QY 900 ATCTGCGGCAAGTTGTGCTGTAACAATATGTCACCTGCTGCTGCTGATGATGATGAT 959
 DB 1126 ATCTGCGGCAAGTTGTGCTGTAACAATATGTCACCTGCTGCTGCTGATGATGATGAT 1185
 QY 960 TGCATTTTGAAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1019
 DB 1186 TGCATTTTGAAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1245
 QY 1020 GCTGTGCAACATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1079
 DB 1246 GCTGTGCAACATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1305
 QY 1080 GCTGTGCAACATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1139
 DB 1306 GCTGTGCAACATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1365
 QY 1140 TGCCTTGAATTAAGAACTCTCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1199
 DB 1366 TGCCTTGAATTAAGAACTCTCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1424
 QY 1200 GGCCTGCACTAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259
 DB 1425 GGCCTGCACTAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1484
 QY 1260 AGACAGGCTGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1319
 DB 1485 AGACAGGCTGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1544
 QY 1320 GTCCTGCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1379
 DB 1545 GTCCTGCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1604
 QY 1380 TCCATGACTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1439

DB 1605 TCCATGACTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1664
 QY 1440 TATTTCTTGGCCACATCTTCTTCTGAGCTTACTATTCATTTGCTTATTTACAAA 1499
 DB 1665 TATTTCTTGGCCACATCTTCTTCTGAGCTTACTATTCATTTGCTTATTTACAAA 1724
 QY 1500 GCAATGAGCCAAAGAAAGAGTAAAGATGAAATATGCAATTTCCCTGCTGCTG 1559
 DB 1725 GCAATGAGCCAAAGAAAGAGTAAAGATGAAATATGCAATTTCCCTGCTGCTGCTG 1784
 QY 1560 TGTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1619
 DB 1785 TGTGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1844
 QY 1620 GCAGAGATGAGAAAGCCAGAGAGTGAAGATGATGCTTCACTGCTGCTGCTGCTGCT 1679
 DB 1845 GCAGAGATGAGAAAGCCAGAGAGTGAAGATGATGCTTCACTGCTGCTGCTGCTGCTG 1904
 QY 1680 CAGCCAGCTTCAATTTGGGCGCAAGGAGGAACTTTTGTGAGAGGCTTGTGCT 1739
 DB 1905 CAGCCAGCTTCAATTTGGGCGCAAGGAGGAACTTTTGTGAGAGGCTTGTGCTGCT 1963
 QY 1740 TGTCAACCAAGCTGAGATGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1799
 DB 1964 TGTCAACCAAGCTGAGATGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2023
 QY 1800 GTTCAAGATGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1859
 DB 2024 GTTCAAGATGATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2083
 QY 1860 GCCCAGCTAATTTTGTATTTTCACTAGAAAGGAGATTTTCACTGCTGCTGCTGCTGCT 1919
 DB 2084 GCCCAGCTAATTTTGTATTTTCACTAGAAAGGAGATTTTCACTGCTGCTGCTGCTGCT 2143
 QY 1920 CTGCAACTCTGACCCGCAAGTATCACCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1979
 DB 2144 CTGCAACTCTGACCCGCAAGTATCACCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2203
 QY 1980 GCGGTGAGCCACCGTCCCGCCGCAAGGAGAACTTTGAGGAGAGCAGAGGCGCTG 2039
 DB 2204 GCGGTGAGCCACCGTCCCGCCGCAAGGAGAACTTTGAGGAGAGCAGAGGCGCTG 2263
 QY 2040 ACATCTCCCTGATGATTTCCCATGACATGCTTATCTGCTGCTGCTGCTGCTGCTGCTGCT 2099
 DB 2264 ACATCTCCCTGATGATTTCCCATGACATGCTTATCTGCTGCTGCTGCTGCTGCTGCTGCT 2323
 QY 2100 CTATGCTGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2159
 DB 2324 CTATGCTGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2383
 QY 2160 CCCCATGAGGAGGCTGAGAGGAGGCTGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2219
 DB 2384 CCCCATGAGGAGGCTGAGAGGAGGCTGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2443
 QY 2220 ACTGATTAAGATTAAGATCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2258
 DB 2444 ACTGATTAAGATTAAGATCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2482

RESULT 4
 LOCUS BC000664 2058 bp mRNA linear PRI 22-APR-2003
 DEFINITION Homo sapiens putative protein similar to neesby (Drosophila), mRNA
 (cDNA clone MGC:1311 IMAGE:334938), complete cds.
 ACCESSION BC000664
 VERSION BC000664.1 GI:12653756
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2058)

AUTHORS	
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Stemann, C.M., Schuler, G.D., Altshuler, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Hsieh, N.K., Hopkin, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Bhat, F., Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldi, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.W., Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.	
TITLE	
human and mouse cDNA sequences	
JOURNAL	
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	
MEDLINE	
23388257	
PUBMED	
12477932	
JOURNAL	
2 (bases 1 to 2058)	
AUTHORS	
Straussberg, R.	
TITLE	
Submitted (15-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	
REMARK	
NIH-MGC Project URL: http://mgc.nci.nih.gov	
COMMENT	
Contact: MGC help desk Email: cgaps-remail.nih.gov Tissue Procurement: ATCC DNA Library Preparation: Rubin Laboratory DNA Library Arrayed by: The I.M.A.G.E. Consortium (LIND) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc.mgc@nih.gov Ahner, N., Ayele, K., Beckstrom-Stenberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Hashiguchi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaesi, R., Maduro, Q.L., Mastello, C., Maekari, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stancik, S., Thomas, P.J., Touchman, J.W., Tsurgou, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.	
FEATURES	
Source	
1. .2058 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:1311 IMAGE:3349388" /tissue_type="Colon, adenocarcinoma" /clone_id="NIH_MGC_15" /lab_host="DH10B-R" /note="Vector: pOT87" 1. .2058 /gene="C3F" /db_xref="LocusID:10162" 177. .1322 /codon_start=1 /product="putative protein similar to nesy" /protein_id="AAH00664.1" /db_xref="GI:12653757"	
gene	
CDS	
FEATURES	
Source	
1. .2058 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:1311 IMAGE:3349388" /tissue_type="Colon, adenocarcinoma" /clone_id="NIH_MGC_15" /lab_host="DH10B-R" /note="Vector: pOT87" 1. .2058 /gene="C3F" /db_xref="LocusID:10162" 177. .1322 /codon_start=1 /product="putative protein similar to nesy" /protein_id="AAH00664.1" /db_xref="GI:12653757"	

/db_xref="LocusID:10162"
 /translation="MGRITPAVLTTPFCROMAYLLAGYYTATGANDIKMTPHCVLT
 KLIGLAVIDPDGKQONSLSSQQRYALRGVSLLEVGFYFAFLVGPFSNMHY
 MLYOGELIDPGKIPNSIIPALKRLSLGFLVLYTLLSPITADLYLTEDYDNHF
 WRCWMIIMGFEVLYKYVTCULVNEGCIILGLFNGEESGKAKMDCAKMWLF
 EIRPPTGIAEFNINTNANWARYI EKRLKFLGNLSGSLFLFALPHGSHGLV
 GROMFLIVIVROARLIQSPISKLAIVIVLQPFYVINOOTTHIMVMSMAFC
 LFTWPKMLKVIYSITFLGHIFLSLFIPLPIYHKMWPKKELKME"
 BASE COUNT 470 a 575 c 457 g 556 t
 ORIGIN
 Query Match 90.0%; Score 2045.4; DB 9; Length 2058;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2046; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	227	GGGTTACCCCTTGGCTTTGTTTATCGGATTAACCTTTCTACAGAGACCTACCTCAT	286
Db	8	GGGTTACCCCTTGGCTTTGTTTATCGGATTAACCTTTCTACAGAGACCTACCTCAT	67
Qy	287	CCACTCTTCCATACCTTTACAGGCTCTCAATGCTTTATTTAACTTTGAAACCAAGCT	346
Db	68	CCACTCTTCCATACCTTTACAGGCTCTCAATGCTTTATTTAACTTTGAAACCAAGCT	127
Qy	347	CTACCACTCCCTGCTGCTGATTTGCTTCACTTCCTCATCTTCGACATAGGCGGCAC	406
Db	128	CTACCACTCCCTGCTGCTGATTTGCTTCACTTCCTCATCTTCGACATAGGCGGCAC	187
Qy	407	CATCACTGCGCTCTCCTACCTTTTGTCTTCAGATGAGCTTCACTTGGCTGATATGA	466
Db	188	CATCACTGCGCTCTCCTACCTTTTGTCTTCAGATGAGCTTCACTTGGCTGATATGA	247
Qy	467	TTACATGTCGACCGGCACTAGATATCAAGTGAACAAGCCATTTGTTTGACTTT	526
Db	248	TTACATGTCGACCGGCACTAGATATCAAGTGAACAAGCCATTTGTTTGACTTT	307
Qy	527	GAAGCTGATGTTGGTGGCTGTGACTACTTTCAGAGGAGGAAAGATCAAACTCTTGC	586
Db	308	GAAGCTGATGTTGGTGGCTGTGACTACTTTCAGAGGAGGAAAGATCAAACTCTTGC	367
Qy	587	CTCTGAGCAACAGAAATATCCATACGATGCTGCTTCTTCTGCTGGAAGTGGCTTT	646
Db	368	CTCTGAGCAACAGAAATATCCATACGATGCTGCTTCTTCTGCTGGAAGTGGCTTT	427
Qy	647	CTCTTACTTCTATGAGGCGCTTCTGTTAGGCGCCGCTTCAATTAATCACTACATGA	706
Db	428	CTCTTACTTCTATGAGGCGCTTCTGTTAGGCGCCGCTTCAATTAATCACTACATGA	487
Qy	707	GCTGTGACAGGAGAGCTGATGACATACAGAGAAATACCAACAGATCTTCTGCG	766
Db	488	GCTGTGACAGGAGAGCTGATGACATACAGAGAAATACCAACAGATCTTCTGCG	547
Qy	767	TCTCAAGCGCTGATGCTGCGCTTCTTCTACTAGTGGGCTACACTGTCAGCCCCA	826
Db	548	TCTCAAGCGCTGATGCTGCGCTTCTTCTACTAGTGGGCTACACTGTCAGCCCCA	607
Qy	827	CATCAAGAAAGTATCTCTCTCACTAAGACTATGACCAACCCCTTGCTGCGCTG	886
Db	608	CATCAAGAAAGTATCTCTCTCACTAAGACTATGACCAACCCCTTGCTGCGCTG	667
Qy	887	CATGTACATGCTGATCTGGGGCAAGTTTGCTGTCAAAATATGTCACCTGTTGGCTGT	946
Db	668	CATGTACATGCTGATCTGGGGCAAGTTTGCTGTCAAAATATGTCACCTGTTGGCTGT	727
Qy	947	CACAGAAGAGATATGATTTTGAAGGCGCTGCGCTTCAATGCTTTGAAGAAAGGCA	1006
Db	728	CACAGAAGAGATATGATTTTGAAGGCGCTGCGCTTCAATGCTTTGAAGAAAGGCA	787
Qy	1007	GGCAAGTGGAGTCTGTGCGCAACATGAAGTGTGGCTTTGAAGAAAGGCGCTT	1066
Db	788	GGCAAGTGGAGTCTGTGCGCAACATGAAGTGTGGCTTTGAAGAAAGGCGCTT	847
Qy	1067	CATGCGACCATGCTCATTTCAATCAACCAACGCGCTGGGTGGCGCTCATATCTT	1126

Db 848 CACTGGACCAATGCTCTCATTCACATCAACCAACGCGCTGGTGCCGCTACATCTT 907
 Qy 1127 CAAAGACTCAAGTTCCTTGAAATTAAGAACTCTCAGGGCTCTCGTGTATCTCT 1186
 Db 908 CAAAGACTCAAGTTCCTTGAAATTAAGAACTCTCAGGGCTCTCGTGTATCTCT 967
 Qy 1187 GGGCTCTGGCAAGGCTCTGAGGATACCTGCTGCTTCAGATGAATTCCTCAT 1246
 Db 968 GGGCTCTGGCAAGGCTCTGAGGATACCTGCTGCTTCAGATGAATTCCTCAT 1027
 Qy 1247 TGTATTTGGAAGCAAGGCTGCGGCTCATTCAGAGAGCCCACTTGACAGCT 1306
 Db 1028 TGTATTTGGAAGCAAGGCTGCGGCTCATTCAGAGAGCCCACTTGACAGCT 1087
 Qy 1307 GGGCCGCAATTAATGCTCTCAGCCCTTCTACTATTGTCAGCAAGACCATCACTGGCT 1366
 Db 1088 GGGCCGCAATTAATGCTCTCAGCCCTTCTACTATTGTCAGCAAGACCATCACTGGCT 1147
 Qy 1367 CTTTCAATGGGTATCTCCAGTACTGCTTCTGCTCTTCACTGGGACAAATGGCTTAAGCT 1426
 Db 1148 CTTTCAATGGGTATCTCCAGTACTGCTTCTGCTCTTCACTGGGACAAATGGCTTAAGCT 1207
 Qy 1427 GATATTAATCATATTTCTCTTGCCCAATCTTCTCTGAGCTTACTATTCATATTTGCC 1486
 Db 1208 GATATTAATCATATTTCTCTTGCCCAATCTTCTCTGAGCTTACTATTCATATTTGCC 1267
 Qy 1487 TATATTTCAAGCAATGATGTCAGAGAGAGAGATTAAGAGATGAATTAATTCAT 1546
 Db 1268 TATATTTCAAGCAATGATGTCAGAGAGAGAGATTAAGAGATGAATTAATTCAT 1327
 Qy 1547 TTCCCTGTGAGCTGTGCGGAGCTGTGTGAGAACTACTCTCTCTCTTTTCAAGCACT 1606
 Db 1328 TTCCCTGTGAGCTGTGCGGAGCTGTGTGAGAACTACTCTCTCTCTTTTCAAGCACT 1387
 Qy 1607 CTTTGGCCCAAGCAAGATGAAGAAACCAAGGAGGTGAGAGATCTCTTCCAGCT 1666
 Db 1388 CTTTGGCCCAAGCAAGATGAAGAAACCAAGGAGGTGAGAGATCTCTTCCAGCT 1447
 Qy 1667 GTGCTCTGCTGCGCAAGCTCTTCAATTTGGGGCCCAAGGGGAACTTTTGTGGAGA 1726
 Db 1448 GTGCTCTGCTGCGCAAGCTCTTCAATTTGGGGCCCAAGGGGAACTTTTGTGGAGA 1507
 Qy 1727 AGGCTCTTGTGTTGTCAACCAAGCTGAGATGAGTGGGAGATCTCAGCTCAACGCAAC 1786
 Db 1508 AGGCTCTTGTGTTGTCAACCAAGCTGAGATGAGTGGGAGATCTCAGCTCAACGCAAC 1567
 Qy 1787 CTCACCTCTGAGGTTCAAGTATTTCTGCTCTCAGCTCCCAAGTACTGGGAATACA 1846
 Db 1568 CTCACCTCTGAGGTTCAAGTATTTCTGCTCTCAGCTCCCAAGTACTGGGAATACA 1627
 Qy 1847 GGCAGGCCACCAATGCCAGCTAATTTTGTATTTTCAAGTGAAGAGGGATTTTCAACAGCT 1906
 Db 1628 GGCAGGCCACCAATGCCAGCTAATTTTGTATTTTCAAGTGAAGAGGGATTTTCAACAGCT 1687
 Qy 1907 TGGCAGGCTGTGTTCTGAACTCTCTGACCGCAAGTATCCACCGCTCTCCGCTCCCAAG 1966
 Db 1688 TGGCAGGCTGTGTTCTGAACTCTCTGACCGCAAGTATCCACCGCTCTCCGCTCCCAAG 1747
 Qy 1967 TGGTGGGATTAAGGCTGAGCAAGCTGTCGCGGCCCAAGAGGGGAACTCTTGTGGAGG 2026
 Db 1748 TGGTGGGATTAAGGCTGAGCAAGCTGTCGCGGCCCAAGAGGGGAACTCTTGTGGAGG 1807
 Qy 2027 AGCAGAGGGGCTCACATCTCCCTCTGATTTCCGCCCATGCAATGCTTATCTCTCCCA 2086
 Db 1808 AGCAGAGGGGCTCACATCTCCCTCTGATTTCCGCCCATGCAATGCTTATCTCTCCCA 1867
 Qy 2087 TCTAGCAGGAATCTATTTGTTTCTTCTGCAATTTACTATGATTTGTATGTGCGG 2146
 Db 1868 TCTAGCAGGAATCTATTTGTTTCTTCTGCAATTTACTATGATTTGTATGTGCGG 1927
 Qy 2147 TACACCAACCCCCCATGAGGGGGGTGAGAGGGGTGCAAGGCGCTGCTCTCACTT 2206
 Db 1928 TACACCAACCCCCCATGAGGGGGGTGAGAGGGGTGCAAGGCGCTGCTCTCACTT 1987

Qy 2207 TTTCTACCTGGAACTGATTAGATTAATACCTCTGTTGTTCAGTTTTCAAAAAA 2266
 Db 1988 TTTCTACCTGGAACTGATTAGATTAATACCTCTGTTGTTCAGTTTTCAAAAAA 2047
 Qy 2267 AAAAAA 2273
 Db 2048 AAAAAA 2054
 RESULT 5
 AX311926 1842 bp DNA linear PAT 09-JAN-2002
 LOCUS Sequence 2435 from Patent WO0194629.
 DEFINITION AX311926
 ACCESSION AX311926
 VERSION AX311926.1 GI:18122560
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS
 1 Young, P.E., Augustus, M., Carter, K.C., Edner, R., Endress, G.,
 Horzigan, S., Soppet, D.R. and Weaver, Z.
 Cancer gene determination and therapeutic screening using signature
 gene sets
 TITLE
 JOURNAL
 Patent: WO 0194629-A 2435 13-DEC-2001:
 Avalon Pharmaceuticals (US)
 FEATURES
 source
 1..1842
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 413 a 525 c 413 g 491 t
 ORIGIN
 Query Match 81.0%; Score 1842; DB 6; Length 1842;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 279 TACTCATTCACCTTTTCCATACCTTTACAGGCTCTCAATGCTTATTTTAACCTTTGA 338
 Db 1 TACTCATTCACCTTTTCCATACCTTTACAGGCTCTCAATGCTTATTTTAACCTTTGA 60
 Qy 339 AACCAAGCTACCACTCCCTGCTGTGATTTGCTTCAAGTCCATCTCGACTAATG 398
 Db 61 AACCAAGCTACCACTCCCTGCTGTGATTTGCTTCAAGTCCATCTCGACTAATG 120
 Qy 399 GGCAGCCACATCACTGCGGCTCTCACTACCTTTTCTTCAGATGGCTTACCTTGCT 458
 Db 121 GGCAGCCACATCACTGCGGCTCTCACTACCTTTTCTTCAGATGGCTTACCTTGCT 180
 Qy 459 GGATACATTAACCTGACCGGCAACTACGATATCAAGTGAAGCAATGCATTTGT 518
 Db 181 GGATACATTAACCTGACCGGCAACTACGATATCAAGTGAAGCAATGCATTTGT 240
 Qy 519 CTGACTTTGAAGCTATTTGCTTGTGCTGCTGCTTGAATCTTGAAGGAGGAAATAGAT 578
 Db 241 CTGACTTTGAAGCTATTTGCTTGTGCTGCTGCTTGAATCTTGAAGGAGGAAATAGAT 300
 Qy 579 TCTCTGCTCTGACACAGAAATATATCATAGTGTGTTCTTCCCTGCTGAAGTT 638
 Db 301 TCTCTGCTCTGACACAGAAATATATCATAGTGTGTTCTTCCCTGCTGAAGTT 360
 Qy 639 GCTGTTTCTCTTACTTATATGAGGCTTCTTGTGAGGCGCCAGTCTCAATGATC 698
 Db 361 GCTGTTTCTCTTACTTATATGAGGCTTCTTGTGAGGCGCCAGTCTCAATGATC 420
 Qy 699 TACATGAAGCTGTGACAGGAGAGCTGATTCATACCAAGAAATATCCAAACGATC 758
 Db 421 TACATGAAGCTGTGACAGGAGAGCTGATTCATACCAAGAAATATCCAAACGATC 480
 Qy 759 ATTCTGCTTCAAGCGCTGATGCTGAGGCTTTTCTACTAGTGGCTACACATGCTC 818

Db 481 ATTCTGCTCTCAAGGCGCTGAGTCTGGGCTTTTCTACCTAGTGGCTTACACATGCTC 540
 Qy 819 AGCCCCACATCAAGAGACTATCTCTCATCTGAAGACTATGACAACACCCCTTCTG 878
 Db 541 AGCCCCACATCAAGAGACTATCTCTCATCTGAAGACTATGACAACACCCCTTCTG 600
 Qy 879 TTCCGCTGATGTAATGCTGATCTGAGGAGAGTTTGTCTGTACAAATATGTCACCTGT 938
 Db 601 TTCCGCTGATGTAATGCTGATCTGAGGAGAGTTTGTCTGTACAAATATGTCACCTGT 660
 Qy 939 TGGCTGCTACAAGAGAGATGATCTTTTGAAGGCTGGGCTTCAATGGCTTTGAAGA 998
 Db 661 TGGCTGCTACAAGAGAGATGATCTTTTGAAGGCTGGGCTTCAATGGCTTTGAAGA 720
 Qy 999 AAGGCAAGGCAAGAGAGAGTGGAGTGGCTGTGCAATGAGTGGCTTTTGAAGCAAA 1058
 Db 721 AAGGCAAGGCAAGAGAGAGTGGAGTGGCTGTGCAATGAGTGGCTTTTGAAGCAAA 780
 Qy 1059 CCCCCTTCACTGGCAACCATGCTCTCATTCACATCAACCAACGCTGGGCTGGCC 1118
 Db 781 CCCCCTTCACTGGCAACCATGCTCTCATTCACATCAACCAACGCTGGGCTGGCC 840
 Qy 1119 TACATCTTCAAGAGACTCAAGTCTTGAATTAAGAACTCTCAAGGCTCTGCTG 1178
 Db 841 TACATCTTCAAGAGACTCAAGTCTTGAATTAAGAACTCTCTCAAGGCTCTGCTG 900
 Qy 1179 CTATCTGCTGCTCTGCGACGCTGCACTGAGATACCTGCTGCTCTGAGATGAA 1238
 Db 901 CTATCTGCTGCTCTGCGACGCTGCACTGAGATACCTGCTGCTCTGAGATGAA 960
 Qy 1239 TTCTCTATGTTATTTGTGAAGAAGAGGCTGCCAGGCTCACTCAAGAGAGAGAGAG 1298
 Db 961 TTCTCTATGTTATTTGTGAAGAAGAGGCTGCCAGGCTCACTCAAGAGAGAGAGAG 1020
 Qy 1299 AGAAGCTGGCGCGCATTAATGCTCTGCTGACGCTCTCACTATTTGGGCAAGAGAG 1358
 Db 1021 AGAAGCTGGCGCGCATTAATGCTCTGCTGACGCTCTCACTATTTGGGCAAGAGAG 1080
 Qy 1359 CACTGGCTCTTCAATGGGTTACTCCATGACTGCTCTGCTCTTCAAGTGGGAAGAA 1418
 Db 1081 CACTGGCTCTTCAATGGGTTACTCCATGACTGCTCTGCTCTTCAAGTGGGAAGAA 1140
 Qy 1419 CTTAAGCTGTAAATTCATCTATTTCTTGGGCAATCTTCTTCTGAGCTTACTATTC 1478
 Db 1141 CTTAAGCTGTAAATTCATCTATTTCTTGGGCAATCTTCTTCTGAGCTTACTATTC 1200
 Qy 1479 ATATGCTGTAAATTCATCAAGAGAGAGTGGCAAGAGAGAGAGAGAGAGAGAGAG 1538
 Db 1201 ATATGCTGTAAATTCATCAAGAGAGAGTGGCAAGAGAGAGAGAGAGAGAGAGAG 1260
 Qy 1539 TAATTCATTTCCCTGCTGCTGCTGCGGAGCTGGTGCAGAACTACTGCTCTCTTTC 1598
 Db 1261 TAATTCATTTCCCTGCTGCTGCTGCGGAGCTGGTGCAGAACTACTGCTCTCTTTC 1320
 Qy 1599 ACAGCACTCTTTTGGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1658
 Db 1321 ACAGCACTCTTTTGGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 Qy 1659 TTCCAGCTGCTGCTGCTGCTGCGAGCAAGTCTTCAATTTGGGAGCCAAAGAGAGAG 1718
 Db 1381 TTCCAGCTGCTGCTGCTGCTGCGAGCAAGTCTTCAATTTGGGAGCCAAAGAGAGAG 1440
 Qy 1719 TTGGAGAGAGAGAGCTTGTCTTGTCAACCAAGCTGAAGTGAAGTGGGAGAGTCTCAG 1778
 Db 1441 TTGGAGAGAGAGAGCTTGTCTTGTCAACCAAGCTGAAGTGAAGTGGGAGAGTCTCAG 1500
 Qy 1779 ACCGCAAGCTTCAAGCTCTTGGGTTCAAGTATTTTCTGCTCAAGCTCTCCAGTACTG 1838
 Db 1501 ACCGCAAGCTTCAAGCTCTTGGGTTCAAGTATTTTCTGCTCAAGCTCTCCAGTACTG 1560
 Qy 1839 GGAATTAAGGAG 1898

Db 1561 GGAATTAAGGAG 1620
 Qy 1899 CACCAAGTGGGCGAGGCTGGTGTCCGAATCTCTGACCGCAAGATTCACCCGCTCCGC 1958
 Db 1621 CACCAAGTGGGCGAGGCTGGTGTCCGAATCTCTGACCGCAAGATTCACCCGCTCCGC 1680
 Qy 1959 TCCCAAGTGGGCGAGGCTGGTGTCCGAATCTCTGACCGCAAGATTCACCCGCTCCGC 2018
 Db 1681 TCCCAAGTGGGCGAGGCTGGTGTCCGAATCTCTGACCGCAAGATTCACCCGCTCCGC 1740
 Qy 2019 GTGGAG 2078
 Db 1741 GTGGAG 1800
 Qy 2079 TCTCCCATCTAGCCAGAGATCTATTTGTTTCTTCTG 2120
 Db 1801 TCTCCCATCTAGCCAGAGATCTATTTGTTTCTTCTG 1842
 RESULT 6
 AX410769
 LOCUS AX410769 1842 bp DNA linear PAT 14-JUN-2002
 DEFINITION Sequence 3416 from Patent WO0229103.
 ACCESSION AX410769
 VERSION AX410769.1 GI:21443474
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1. Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J. G.
 Gene expression profiles in liver cancer
 Patent: WO 0229103-A 3416 11-APR-2002;
 JOURNAL GENE LOGIC INC (US)
 FEATURES
 source
 1..1842
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /note="EMBL/Genbank Accession No. U72515"
 BASE COUNT 413 a 525 c 413 g 491 t
 ORIGIN
 Query Match 81.0%; Score 1842; DB 6; Length 1842;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 279 TACCTATCAACCTCTTCCATACCTTACAGGCTCTCAATGCTTATTTAATTGGA 338
 Db 1 TACCTATCAACCTCTTCCATACCTTACAGGCTCTCAATGCTTATTTAATTGGA 60
 Qy 339 AACCACTTCAACCACTCCCTGCTGTGTATTTGCTTCAAGTCTCTCATCTTGA 398
 Db 61 AACCACTTCAACCACTCCCTGCTGTGTATTTGCTTCAAGTCTCTCATCTTGA 120
 Qy 399 GGCCTGCAATCACTGCGCTCTCTCACTACTTTTGTCTCAAGTGGCTTCTGCGCT 458
 Db 121 GGCCTGCAATCACTGCGCTCTCTCACTACTTTTGTCTCAAGTGGCTTCTGCGCT 180
 Qy 459 GGATATTAATCACTGCAACCGGCACTAGATATCAAGTGAAGTGAAGTGAAGTGTGT 518
 Db 181 GGATATTAATCACTGCAACCGGCACTAGATATCAAGTGAAGTGAAGTGAAGTGTGT 240
 Qy 519 CTGACTTTGAAGTGAATGTTGCTTGTGCTTGAAGTGAAGTGAAGTGAAGTGAAGT 578
 Db 241 CTGACTTTGAAGTGAATGTTGCTTGTGCTTGAAGTGAAGTGAAGTGAAGTGAAGT 300
 Qy 579 TCCCTTGTCTCTGAGCAAGAAATATGCAATCGTGTGCTTCTCCCTGCGAAGT 638
 Db 301 TCCCTTGTCTCTGAGCAAGAAATATGCAATCGTGTGCTTCTCCCTGCGAAGT 360
 Qy 639 GCTGTTTCTCTACTTATATGAGGCTTCTTGTGAGGCGCCAGTTCTCATGAATCAC 698


```

|||||
361 GGTGGTTTCTCTACTTCTATGGGGCTTCTTGTTGGGGCCCGAGTTCTCAATGAATAC 420
|||
699 TACATGAAGCTGTGAGGAGAGCTGATGACATACCGAGAAAGATACCAACAGCATC 758
|||
421 TACATGAAGCTGTGAGGAGAGCTGATGACATACCGAGAAAGATACCAACAGCATC 480
|||
759 ATTCTGCTCTCAAGGCTGAGTCTGGGCTTTTCTACCTAGTGGGCTACACCTGTC 818
|||
481 ATTCTGCTCTCAAGGCTGAGTCTGGGCTTTTCTACCTAGTGGGCTACACCTGTC 540
|||
819 AGCCCCCATCATCAGAGAACTATCTCTCACTGAAGACTATGACCAACCCCTTCTG 878
|||
541 AGCCCCCATCATCAGAGAACTATCTCTCACTGAAGACTATGACCAACCCCTTCTG 600
|||
879 TTCCCGTGATGACATGCTGATCTGGGGCAAGTTTGCTGTAACAATAATGACCTGT 938
|||
601 TTCCCGTGATGACATGCTGATCTGGGGCAAGTTTGCTGTAACAATAATGACCTGT 660
|||
939 TGGCTGTGACAGAGAGATATGATTTTGAACGGGCTTCAATGGCTTTGAAGAA 998
|||
661 TGGCTGTGACAGAGAGATATGATTTTGAACGGGCTTCAATGGCTTTGAAGAA 720
|||
999 AAGGCGAAGCAAGTGGAGTGTGCTGCTCAACATGAGGTGTGCTCTTTGAACAAAC 1058
|||
721 AAGGCGAAGCAAGTGGAGTGTGCTGCTCAACATGAGGTGTGCTCTTTGAACAAAC 780
|||
1059 CCCCCTTCACTGGGACCAATGCTGATTCACATACCAACCGCTGGTGGCCCGC 1118
|||
781 CCCCCTTCACTGGGACCAATGCTGATTCACATACCAACCGCTGGTGGCCCGC 840
|||
1119 TACATCTTCAAGCACTCAAGTCTCTTGAATAAAGAACTCTGAGGATCTCTGTTG 1178
|||
841 TACATCTTCAAGCACTCAAGTCTCTTGAATAAAGAACTCTGAGGATCTCTGTTG 900
|||
1179 CTATTCCTGGCCCTTGGGACCGGCTGCACTCAGATACCTGGTCTGCTTCAGATGAA 1238
|||
901 CTATTCCTGGCCCTTGGGACCGGCTGCACTCAGATACCTGGTCTGCTTCAGATGAA 960
|||
1239 TTCCCTATGTTATTTGTGAAAAGACAGGCTGCCAGGCTCATTCAGAGAGCCCACTCG 1298
|||
961 TTCCCTATGTTATTTGTGAAAAGACAGGCTGCCAGGCTCATTCAGAGAGCCCACTCG 1020
|||
1299 ACCAAGCTGGCCGCACTACTGCTCTCCAGCCCTTCTACTATTTGGTGAACAGACATC 1358
|||
1021 ACCAAGCTGGCCGCACTACTGCTCTCCAGCCCTTCTACTATTTGGTGAACAGACATC 1080
|||
1359 CACTGCTCTTCAATGGGTTACTTCATGACTGCTTCTGCTCTTCACTGAGCAATATG 1418
|||
1081 CACTGCTCTTCAATGGGTTACTTCATGACTGCTTCTGCTCTTCACTGAGCAATATG 1140
|||
1419 CTTAAGGTATTAATTCATCTATTTCTTGGGCAACATCTTCTCTGAGCTACTATTC 1478
|||
1141 CTTAAGGTATTAATTCATCTATTTCTTGGGCAACATCTTCTCTGAGCTACTATTC 1200
|||
1479 AATATGCTTATATTCAGAAAGCATGTGTGCCAAGAAAGAAAGTAAAGAAATGAA 1538
|||
1201 AATATGCTTATATTCAGAAAGCATGTGTGCCAAGAAAGAAAGTAAAGAAATGAA 1260
|||
1539 TAATCATTTCCCTGTGGCTGTGCGGGAAGTGTGCAAGAACTACTGCTCTCCCTTTTC 1598
|||
1261 TAATCATTTCCCTGTGGCTGTGCGGGAAGTGTGCAAGAACTACTGCTCTCCCTTTTC 1320
|||
1599 AAGAGCATCTCTTGGCCCGCAGAGAGAGATGAAAAGCCAGGAGGTGAAGATGCATGC 1658
|||
1321 AAGAGCATCTCTTGGCCCGCAGAGAGAGATGAAAAGCCAGGAGGTGAAGATGCATGC 1380
|||
1659 TTCCAGCTGTGCTCTGCTGCCAGCAAGTCTTCAATTTGGGGCCAAAGGGGAACTTTT 1718
|||
1381 TTCCAGCTGTGCTCTGCTGCCAGCAAGTCTTCAATTTGGGGCCAAAGGGGAACTTTT 1440
|||
1719 TTGGAGAAAGGCGCTTGTGCTTTGTCAACCAAGCTGGAATGCACTGGCGGGAATCTCAGCTC 1778
|||||

```

```

Db 1441 TTGGAGAAAGCGCTCTTGTGTCACCCAGCTGGAATGAGTGGCGGATCTCAGCTC 1500
|||
Qy 1779 ACCGAACTCCACCTCCTGGGTTCAAGTAAATTTCTGCTCAGCTCCCAAGATGCTG 1838
|||
Db 1501 ACCGAACTCCACCTCCTGGGTTCAAGTAAATTTCTGCTCAGCTCCCAAGATGCTG 1560
|||
Qy 1839 GGAATACAGGACGCGCACATGCCAGCTAAATTTTGTATTTTCACTAGAAAACGGATTT 1898
|||
Db 1561 GGAATACAGGACGCGCACATGCCAGCTAAATTTTGTATTTTCACTAGAAAACGGATTT 1620
|||
Qy 1899 CACCACTTTGGCCAGCTGTGCTGCACTCTGACCGCAAGTATCCACCGCTCCGCC 1958
|||
Db 1621 CACCACTTTGGCCAGCTGTGCTGCACTCTGACCGCAAGTATCCACCGCTCCGCC 1680
|||
Qy 1959 TCCCAAGTGTGGGATTAACAGCGCTGAGCCACCGTCCCGGCCCAAGGGGAACTCTT 2018
|||
Db 1681 TCCCAAGTGTGGGATTAACAGCGCTGAGCCACCGTCCCGGCCCAAGGGGAACTCTT 1740
|||
Qy 2019 GTGGAGAGAGAGAGGAGCTCACATCTCCCTGATTTCCCATGACATTTGCTTATC 2078
|||
Db 1741 GTGGAGAGAGAGAGGAGCTCACATCTCCCTGATTTCCCATGACATTTGCTTATC 1800
|||
Qy 2079 TCTCCCATCTAGCCAGGAATCTATGTGTTTCTTCTGCC 2120
|||
Db 1801 TCTCCCATCTAGCCAGGAATCTATGTGTTTCTTCTGCC 1842
|||
RESULT 7
HSU72515 1842 bp mRNA linear PRI 24-JUL-1997
LOCUS HSU72515
DEFINITION Human C3f mRNA, complete cds.
ACCESSION U72515
VERSION U72515.1 GI:1673519
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Ansari-Lari, M.A., Shen, Y., Muzny, D.M., Lee, W. and Gibbs, R.A.
TITLE Large-scale sequencing in human chromosome 12p13: experimental and
JOURNAL computational gene structure determination
GENEID Genome Res. 7 (3), 268-280 (1997)
MEDLINE 97228904
PUBMED 9074930
REFERENCE
AUTHORS Ansari-Lari, M.A., Shen, Y., Muzny, D.M., Lee, W. and Gibbs, R.A.
TITLE Direct Submision
JOURNAL Submitted (24-SEP-1996) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
FEATURES
source
1..1842
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12p13"
1..1842
/gene="C3f"
1..1842
/misc_feature
1..1842
/gene="C3f"
118..1263
/gene="C3f"
CDS
118..1263
/gene="C3f"
/note="similar to ESTs with GenBank Accession Numbers
H45806 and H45837; similar to S. cerevisiae ORF YOR175c
encoded by GenBank Accession Number Z75083; see
corresponding genomic sequence in GenBank Accession Number
U72506"
/codon_start=1
/product="C3f"
/protein_id="AAC51640.1"
/db_xref="GI:1673520"
/translation="MGRTITAVLTFQFQAVLLAGYVYTAATGNVDIKMTBPHCVLT

```

KLIGLADVYEDGKDNLSSEBQKXAIRGVSLAEVAGSYFYAGFLVGPORSNMHY
MKLVQSELDIDPEKIPNSITIPALRISLSGLFVLVGTTLSTBYLITBEDNDHP
WERCMYMLWIKFVLYKYVTCWLVEGVCLLGLGNGEERKEKRAKANNKWLIF
ENRPRFTIASFNINAMVARYIFRKLFLNKL.SQSLSLFLALHMDHSGSLV
CFQMEFLIVEROARLIOESPILSKLAIVLOPFPYVLOOTIHLIMVSGYSMTAF
LFTWDMKLVKYSIYFLGHIFFLSLFIPIYIKHAWVPRKXKXME"

BASE COUNT 413 a 525 c 413 g 491 t
Query Match 81.0%; Score 1842; DB 9; Length 1842;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 TACCTCATCACTCTTCCATACCTTTACAGGCTCTCAATGGCTATTTAACTTTGA 338
DB 1 TACCTCATCACTCTTCCATACCTTTACAGGCTCTCAATGGCTATTTAACTTTGA 60
QY 339 AACCAAGCTCAACCACTCCCTGCTGTGTAATGGCTTCAATCCCTCACTCAATG 398
DB 61 AACCAAGCTCAACCACTCCCTGCTGTGTAATGGCTTCAATCCCTCACTCAATG 120
QY 399 GGGCGACATCACTGCGCTCTCACTACCTTTTGTCCAGATGGCTTCACTTGGCT 458
DB 121 GGGCGACATCACTGCGCTCTCACTACCTTTTGTCCAGATGGCTTCACTTGGCT 180
QY 459 GGAATCAATTTACATGCGCACTGCGCACTACGATATCAAGTGGACATTTGTT 518
DB 181 GGAATCAATTTACATGCGCACTGCGCACTACGATATCAAGTGGACATTTGTT 240
QY 519 CTACCTTTGAAGCTGATTTGTTGGCTGTGTAATGGCTTCAATGGCTTCAATG 578
DB 241 CTACCTTTGAAGCTGATTTGTTGGCTGTGTAATGGCTTCAATGGCTTCAATG 300
QY 579 TCCTTGTCTCTGAGCAAGAAATATGCAATGCTGTGTTCTTCCCTGCTGGAAT 638
DB 301 TCCTTGTCTCTGAGCAAGAAATATGCAATGCTGTGTTCTTCCCTGCTGGAAT 360
QY 639 GCTGTGTTCTCTCACTTTCTATGCGGCTTCTTGTGAGGCGCCCACTTCAATGATCAC 658
DB 361 GCTGTGTTCTCTCACTTTCTATGCGGCTTCTTGTGAGGCGCCCACTTCAATGATCAC 420
QY 699 TACATGAAGCTGTGAGGAGAGCTGATTTGATCACTACAGAAAGATACCAACAGCATC 758
DB 421 TACATGAAGCTGTGAGGAGAGCTGATTTGATCACTACAGAAAGATACCAACAGCATC 480
QY 759 ATTCTGCTCTCAAGCGCTGAGTCTGGGCTTTTCTACCTAGTGGCTTCACTGCTC 818
DB 481 ATTCTGCTCTCAAGCGCTGAGTCTGGGCTTTTCTACCTAGTGGCTTCACTGCTC 540
QY 819 AGCCCCCACTACAGAAAGCTATCTCTCACTGAAAGATATGACCAACCCCTTCTGG 878
DB 541 AGCCCCCACTACAGAAAGCTATCTCTCACTGAAAGATATGACCAACCCCTTCTGG 600
QY 879 TTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 938
DB 601 TTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 939 TGGCTGTCTACAGAAAGATGATTTTGAAGGCTGAGCTTCAATGGCTTTGAAGAA 998
DB 661 TGGCTGTCTACAGAAAGATGATTTTGAAGGCTGAGCTTCAATGGCTTTGAAGAA 720
QY 999 AAGGCAAGGCAAGTGGATGCTGTGCGCAATGAAAGTGTGCTTTTGAAGCAAC 1058
DB 721 AAGGCAAGGCAAGTGGATGCTGTGCGCAATGAAAGTGTGCTTTTGAAGCAAC 780
QY 1059 CCCGCTTCACTGAGCACTATGCTCATTTCAATCAACCAACGCTGGGTGGCCGC 1118
DB 781 CCCGCTTCACTGAGCACTATGCTCATTTCAATCAACCAACGCTGGGTGGCCGC 840
QY 1119 TACATCTTCAACGACTCAAGTTCTTGAAGTAAAGACTCTCAAGGCTCTCTGTT 1178
DB 841 TACATCTTCAACGACTCAAGTTCTTGAAGTAAAGACTCTCAAGGCTCTCTGTT 900

QY 1179 CTATTCCTGGCCCTCTGAGACGGCCCTGACCTGAGATACCTGGTCTGCTTCCAGATGAA 1238
DB 901 CTATTCCTGGCCCTCTGAGACGGCCCTGACCTGAGATACCTGGTCTGCTTCCAGATGAA 960
QY 1239 TTCCTCATGTTATTTGTGAAAGAGACAGCTGCGCAGGCTCATTTCAAGAGAGCCCACTTG 1298
DB 961 TTCCTCATGTTATTTGTGAAAGAGACAGCTGCGCAGGCTCATTTCAAGAGAGCCCACTTG 1020
QY 1299 AGCAAGCTGGCCGCTTACTGTCTCTGACGCCCTTCTACTATTTGGTGCACAGACCATC 1358
DB 1021 AGCAAGCTGGCCGCTTACTGTCTCTGACGCCCTTCTACTATTTGGTGCACAGACCATC 1080
QY 1359 CACTGCTCTTCAATGGCTTACTGCTCAATGATGCTGCTGCTGCTTCAAGTGGGCAATATG 1418
DB 1081 CACTGCTCTTCAATGGCTTACTGCTCAATGATGCTGCTGCTGCTTCAAGTGGGCAATATG 1140
QY 1419 CTTAAGGTATTAATTCATCTATTTCTTGGCCACATCTTCTCTGAGCTTACTATTC 1478
DB 1141 CTTAAGGTATTAATTCATCTATTTCTTGGCCACATCTTCTCTGAGCTTACTATTC 1200
QY 1479 ATATTCCTTATTTTCAAAAGCAATGTTGCCAAGAAAGAGATTAAGAAATGAA 1538
DB 1201 ATATTCCTTATTTTCAAAAGCAATGTTGCCAAGAAAGAGATTAAGAAATGAA 1260
QY 1539 TAATCATTTCCCTGGTGGCTGTGGGAGCTGTGCAAAACTACTGCTTCCCTTTC 1598
DB 1261 TAATCATTTCCCTGGTGGCTGTGGGAGCTGTGCAAAACTACTGCTTCCCTTTC 1320
QY 1599 ACAGCACTCTTCTTGGCCGAGAGAGATGAAAGCCAGGAGGTGAGATTCGATGC 1658
DB 1321 ACAGCACTCTTCTTGGCCGAGAGAGATGAAAGCCAGGAGGTGAGATTCGATGC 1380
QY 1659 TTCAGCTGTGCTGTGCTGCGACGCAAGTCTTATTTGGGCGCAAGGGGAAACTTTT 1718
DB 1381 TTCAGCTGTGCTGTGCTGCGACGCAAGTCTTATTTGGGCGCAAGGGGAAACTTTT 1440
QY 1719 TTTGAGAAAGGCTCTTGTGTTGTGCAACGCTGGAAGTGTGCGGGGATCTGAGCT 1778
DB 1441 TTTGAGAAAGGCTCTTGTGTTGTGCAACGCTGGAAGTGTGCGGGGATCTGAGCT 1500
QY 1779 ACCGCAACTCTCACTCTGAGGTTCAAGTATTTCTGCTCAGCCTCCCAAGTATGCTG 1838
DB 1501 ACCGCAACTCTCACTCTGAGGTTCAAGTATTTCTGCTCAGCCTCCCAAGTATGCTG 1560
QY 1839 GGAATACAGGACGCGCACTATGCTCAATTTTGTATTTTCAATGAAAGGGAAT 1898
DB 1561 GGAATACAGGACGCGCACTATGCTCAATTTTGTATTTTCAATGAAAGGGAAT 1620
QY 1899 CACCAAGTTGGCCAGGCTGTGCTGAACTCTGACCGCAAGTATCACCCGCTCGGC 1958
DB 1621 CACCAAGTTGGCCAGGCTGTGCTGAACTCTGACCGCAAGTATCACCCGCTCGGC 1680
QY 1959 TCCCAAGTGTGGGATTTACAGGCTGAGCACTGTCGCGCCCAAGGGGAAACTCTT 2018
DB 1681 TCCCAAGTGTGGGATTTACAGGCTGAGCACTGTCGCGCCCAAGGGGAAACTCTT 1740
QY 2019 GTGGAGAGAGCAGAGGAGCTCACATCTCCCTCTGATTTCCCATGACATTCCTTATC 2078
DB 1741 GTGGAGAGAGCAGAGGAGCTCACATCTCCCTCTGATTTCCCATGACATTCCTTATC 1800
QY 2079 TCTCCCATCTAGCCAGGAATCTATTGTGTTTCTTCTGCC 2120
DB 1801 TCTCCCATCTAGCCAGGAATCTATTGTGTTTCTTCTGCC 1842

RESULT 8
AX073560
LOCUS AX073560 1461 bp DNA linear PAT 06-FEB-2001
DEFINITION Sequence 12 from Patent WO0104297.
ACCESSION AX073560
VERSION AX073560.1 GI:12709973
KEYWORDS
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1
 Kato, S. and Kimura, T.
 Human proteins having hydrophobic domains and dnas encoding these
 proteins
 Patent: WO 0104297-A 12 18-JAN-2001;
 SAGAMI CHEMICAL RESEARCH CENTER (JP) ; Proteogene Inc. (JP)
 FEATURES
 source 1. 1461
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 326 a 401 c 336 g 398 t
 ORIGIN
 Query Match 64.3%; Score 1461; DB 6; Length 1461;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1461; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 78 ATGGCTCTCTACGCGAGGGGGAAGAGGAGCTGTGTGGCGCTGGCGGGGTTCTGCAAG 137
 DB 1 ATGGCGTCTCTACGCGAGGGGGAAGAGGAGCTGTGTGGCGCTGGCGGGGTTCTGCAAG 60
 QY 138 TCGGGTTTCAGAGAGCTGAGCTTAAACAAGTTGGGAGCGCTGGGGCGGTGAGAAAG 197
 DB 61 TCGGGTTTCAGAGAGCTGAGCTTAAACAAGTTGGGAGCGCTGGGGCGGTGAGAAAG 120
 QY 198 GCGCTGCGGCTGATCATCTTCATCTTCTGGGTTACCCCTTTGCTTTTATCGGCAT 257
 DB 121 GCGCTGCGGCTGATCATCTTCATCTTCTGGGTTACCCCTTTGCTTTTATCGGCAT 180
 QY 258 TACCTTTTCTACAGAGAGACCTACCTACCTACCTCTTCTGATACCTTTACAGGCTCTCA 317
 DB 181 TACCTTTTCTACAGAGAGACCTACCTACCTACCTCTTCTGATACCTTTACAGGCTCTCA 240
 QY 318 ATGCTTATTTTAACTTTGGAAACAGCTCTACACCTCCGCTGCTGATTTGCTTCAAG 377
 DB 241 ATGCTTATTTTAACTTTGGAAACAGCTCTACACCTCCGCTGCTGATTTGCTTCAAG 300
 QY 378 TTCCTCATCTTGAGTAAATGGGCGGACCATCATCTGCGTCTCTCACTACCTTTTGGCTTC 437
 DB 301 TTCCTCATCTTGAGTAAATGGGCGGACCATCATCTGCGTCTCTCACTACCTTTTGGCTTC 360
 QY 438 CAGATGGCTACCTTCTGCTGATATCTATTAACCTGCGACCGGCACTACGATATCAAG 497
 DB 361 CAGATGGCTACCTTCTGCTGATATCTATTAACCTGCGACCGGCACTACGATATCAAG 420
 QY 498 TGGACAATGCGACATGTTGTTCTGACTTTGAAGCTGATTTGGCTTTTGACTACTTT 557
 DB 421 TGGACAATGCGACATGTTGTTCTGACTTTGAAGCTGATTTGGCTTTTGACTACTTT 480
 QY 558 GAGGAGGAGAAAGATCAGATTCCTTGTCTCTGAGCAACAGAAATATGCCATACCTGTGT 617
 DB 481 GAGGAGGAGAAAGATCAGATTCCTTGTCTCTGAGCAACAGAAATATGCCATACCTGTGT 540
 QY 618 GTTCTCTCTCTGCTGAGAGTTGCTGTTTCTCTACTTCTATGAGGCTTTCTTGGTGAAG 677
 DB 541 GTTCTCTCTCTGCTGAGAGTTGCTGTTTCTCTACTTCTATGAGGCTTTCTTGGTGAAG 600
 QY 678 CCCGAGTTTCAATGAATCACTAATGAGCTGTGTGAGGAGAGATGATTTGACATACCA 737
 DB 601 CCCGAGTTTCAATGAATCACTAATGAGCTGTGTGAGGAGAGATGATTTGACATACCA 660
 QY 738 GGAAGATATCAAAACAGCATCATCTCTGCTTCAAGCGCTGAGTCTGGGCTTTTCTAC 797
 DB 661 GGAAGATATCAAAACAGCATCATCTCTGCTTCAAGCGCTGAGTCTGGGCTTTTCTAC 720
 QY 798 CTATGTTGGGCTACACTGCTCAAGCTCCCATATCAAGAGACTATCTCTCACTGAAGAC 857
 DB 721 CTATGTTGGGCTACACTGCTCAAGCTCCCATATCAAGAGACTATCTCTCACTGAAGAC 780

QY 858 TATGACAACCAACCCCTTGTGTTCCGCTGACATGATCATGCTGATCTGGGCGAGTTTGTG 917
 DB 781 TATGACAACCAACCCCTTGTGTTCCGCTGACATGATCATGCTGATCTGGGCGAGTTTGTG 840
 QY 918 CTGTAACAATATGTCACCTGTTGGTGTGTCACAGAGAGATATGCAATTTTGAAGGCGCTG 977
 DB 841 CTGTAACAATATGTCACCTGTTGGTGTGTCACAGAGAGATATGCAATTTTGAAGGCGCTG 900
 QY 978 GGGCTCATGCTTTGGAAGAAAGGCAAGGCAAGGATGAGTGGTGGCCCAATATGAG 1037
 DB 901 GGGCTCATGCTTTGGAAGAAAGGCAAGGCAAGGATGAGTGGTGGCCCAATATGAG 960
 QY 1038 GTGTGGCTTTTGAAGAAACCCCGCTTCACTGSCACCATTTGCTCAATTCATCAAC 1097
 DB 961 GTGTGGCTTTTGAAGAAACCCCGCTTCACTGSCACCATTTGCTCAATTCATCAAC 1020
 QY 1098 ACCAAGCGCTGGGTGGCCCGCTACATCTTCAACAGCATCAAGTTCTTGGAAATTAAGA 1157
 DB 1021 ACCAAGCGCTGGGTGGCCCGCTACATCTTCAACAGCATCAAGTTCTTGGAAATTAAGA 1080
 QY 1158 CTCTCTACAGGCTCTCTGTTGATTTCTGCTGAGCCCTCTGGACGCGCTGCACCTCAGATAC 1217
 DB 1081 CTCTCTACAGGCTCTCTGTTGATTTCTGCTGAGCCCTCTGGACGCGCTGCACCTCAGATAC 1140
 QY 1218 CTGCTCTGCTTCAAGATGAAATTTCTCATTTGTTATTTGAGAAAGACAGGCTCCAGGCTC 1277
 DB 1141 CTGCTCTGCTTCAAGATGAAATTTCTCATTTGTTATTTGAGAAAGACAGGCTCCAGGCTC 1200
 QY 1278 ATTCAAGAGAGCCCAACCCCTGAGCAAGCTGGCGGCATTTACTGTCTCAGGCTTTGAC 1337
 DB 1201 ATTCAAGAGAGCCCAACCCCTGAGCAAGCTGGCGGCATTTACTGTCTCAGGCTTTGAC 1260
 QY 1338 TATTTGGTGCACAACAACCATCCATGCTCTTCAAGGCTTACTCCATGATGCTCTTGGC 1397
 DB 1261 TATTTGGTGCACAACAACCATCCATGCTCTTCAAGGCTTACTCCATGATGCTCTTGGC 1320
 QY 1398 CTCTCAGGTGGGACAATAGGCTTAAGGTGATTAATCATCATTTCTTGGCCACATC 1457
 DB 1321 CTCTCAGGTGGGACAATAGGCTTAAGGTGATTAATCATCATTTCTTGGCCACATC 1380
 QY 1458 TTCTTCTGAGCTTACTATTCATATTTGCTTATTTCAACAAGCAATGTTGCCAAGAAA 1517
 DB 1381 TTCTTCTGAGCTTACTATTCATATTTGCTTATTTCAACAAGCAATGTTGCCAAGAAA 1440
 QY 1518 GAGAAGTTAAGAAGATGGA 1538
 DB 1441 GAGAAGTTAAGAAGATGGA 1461
 RESULT 9
 AK058063 1573 bp mRNA linear PRI 31-OCT-2001
 LOCUS
 DEFINITION Homo sapiens cDNA FLJ25334 f1s, clone TST00701.
 AK058063
 ACCESSION
 VERSION 1 GI:16554077
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1
 Ieshibashi, T., Kanehori, K., Yoshida, M., Matsunabe, S., Ieshida, S.,
 Ono, Y., Horiuchi, T., Hiraoka, S., Murakawa, K., Takiguchi, S.,
 Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ieshida, M.,
 Yamashita, H., Chiba, Y., Suzuki, Y., Hata, H., Nakagawa, K., Mizuno, S.,
 Morinaga, M., Kawamura, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,
 Nishikawa, T., Sugiyama, A., Kawakami, B., Nagai, K., Isogai, T. and
 Sugano, S.
 NEDO human cDNA sequencing project
 TITLE
 JOURNAL
 Unpublished
 2 (bases 1 to 1573)
 REFERENCE
 Sugano, S. and Suzuki, Y.
 Direct Submission

JOURNAL

Submitted (24-OCT-2001) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)

COMMENT

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction and 5'-end one pass sequencing: Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; 3'-end one pass sequencing: RAB; clone selection for full insert sequencing: RAB and Helix Research Institute.

FEATURES

Source

1. 1573
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="TST00701"
/cissue_type="testis"
/clone_id="TST"
/note="cloning vector: pME18SFL3"

BASE COUNT

358 a 436 c 378 g 401 t

ORIGIN

Query Match 61.1%; Score 1388.2; DB 9; Length 1573;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 1408; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

814 TGCTGACCCCAATCAGAGAGACTATCTCTGGAAGACTATGACACACCCCT 873
133 TGTTCATCAGCCAGCACTGAAACCTTCGGGGCAGGGCTGTGCAACACCCCT 192
874 TCTGTTCCGCTGATCATGCTGATCTGGGGCAAGTTTGTCTGTAATAATGTCA 933
193 TCTGTTCCGCTGATCATGCTGATCTGGGGCAAGTTTGTCTGTAATAATGTCA 252
934 CCGTGGCTGGTCAAGAGAGATGATGATTTGAGGGGCTGGGGCTTAAGGGCTTG 993
253 CCGTGGCTGGTCAAGAGAGATGATGATTTGAGGGGCTGGGGCTTAAGGGCTTG 312
994 AAGAAAAGGCAAGCAAGAGAGATGATGATTTGAGGGGCTGGGGCTTAAGGGCTTG 1053
313 AAGAAAAGGCAAGCAAGAGAGATGATGATTTGAGGGGCTGGGGCTTAAGGGCTTG 372
1054 CAAACCCCGCTTCACTGCAACATGCTGCTTCAACATCAACCAAGCGCTGGGTG 1113
373 CAAACCCCGCTTCACTGCAACATGCTGCTTCAACATCAACCAAGCGCTGGGTG 432
1114 CCGGCTGATCTTCAACAGACTCAAGTTCTTGGAAATTAAGAACTCTCAGGGTCT 1173
433 CCGGCTGATCTTCAACAGACTCAAGTTCTTGGAAATTAAGAACTCTCAGGGTCT 492
1174 CGTTGCTATTCCTGGCCCTTGGCAGCGGCTGCACTCAGATACCTGCTGCTTCAGA 1233
493 CGTTGCTATTCCTGGCCCTTGGCAGCGGCTGCACTCAGATACCTGCTGCTTCAGA 552
1234 TGGAAATCTCTGATTTGTTGTAAGAGAGCTGCGAGGCTTCAATCAAGAGGCCCA 1293
553 TGGAAATCTCTGATTTGTTGTAAGAGAGCTGCGAGGCTTCAATCAAGAGGCCCA 612
1294 CCGGAGCAAGCTGGCGGCACTTACGTCCTCCAGCCCTTCTATTTATTTGGTGAACA 1353
613 CCGGAGCAAGCTGGCGGCACTTACGTCCTCCAGCCCTTCTATTTATTTGGTGAACA 672
1354 CCAATCACTGGCTTCTGATGGTTACTGATGATGCTTCTGCTTCAACGTGGACA 1413
673 CCAATCACTGGCTTCTGATGGTTACTGATGATGCTTCTGCTTCAACGTGGACA 732
1414 AATGGCTTAAGGTATTAATCACTGATTTCTTTGGCCACATCTTCTCTGAGCTAC 1473
733 AATGGCTTAAGGTATTAATCACTGATTTCTTTGGCCACATCTTCTCTGAGCTAC 792
1474 TATTCATATTCCTTATTTCAAAAGCAATGGTGCAAGAGAAAGTAAGTAAGAGA 1533

Db 793 TATTCATATTCCTTATTTATTTCAAAAGCAATGGTGCAAGAGAAAGTAAGTAAGAGA 852
1534 TGGAAATCAATTTCCCTGATGGCTGTGGCGGACTGTGTGAGAAATCACTGCTGCC 1593
853 TGGAAATCAATTTCCCTGATGGCTGTGGCGGACTGTGTGAGAAATCACTGCTGCC 912
1594 TTTTCACAGACTCTCTTTCCTTCCAGAGAGAAATGAGAAATGAGAGAGTGAATG 1653
913 TTTTCACAGACTCTCTTTCCTTCCAGAGAGAAATGAGAAATGAGAGAGTGAATG 972
1654 GATGCTTCAGCTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1713
973 GATGCTTCAGCTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1032
1714 TTTTTCAG 1773
1033 TTTTTCAG 1092
1774 AGCTCAGCGCAACCTTCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1833
1093 AGCTCAGCGCAACCTTCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1152
1834 AGCTGGAATPACAGGAG 1893
1153 AGCTGGAATPACAGGAG 1212
1894 GATTTACACAGCTGTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1953
1213 GATTTACACAGCTGTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1272
1954 CCGCTCCCAAG 2013
1273 CCGCTCCCAAG 1332
2014 CTTTGTGGAG 2073
1333 CTTTGTGGAG 1392
2074 TTATCTCTCCCAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2133
1393 TTATCTCTCCCAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1452
2134 TGTGTAATGCGCTGAC 2193
1453 TGTGTAATGCGCTGAC 1512
2194 GCCTGCTCACTTTTCTTCACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2253
1513 GCCTGCTCACTTTTCTTCACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1572
2254 T 2254
1573 T 1573

RESULT 10
AK098090 2128 bp mRNA linear ROD 25-SEP-2002
LOCUS Mus musculus cDNA fis, clone TRACH2004292, highly similar to Human
DEFINITION AK098090
ACCESSION AK098090
VERSION AK098090.1 GI:21758026
KEYWORDS oligo capping, fis (full insert sequence).
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K., Arita, M., Mutsaers, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hito, Y., Saito, K., Nishikawa, T., Kimura, K.,

Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuno, Y., Negai, K. and Isogai, T.

NEDO cDNA sequencing project

Unpublished

2 (bases 1 to 2128)

Isogai, T. and Yamamoto, J.

Direct Submission

Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan

(E-mail: genom@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)

NEDO cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology (RAB) (supported by Japan Key Technology Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing; RAB, HRI, and Biotechnology Center; National Institute of Technology and Evaluation; clone selection for full insert sequencing; HRI and RAB; annotation: HRI and RAB.

Location/Qualifiers

1..2128

location/Qualifiers

source

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="TRACH2004292"

/note="cloning vector: pME18SFL3"

BASE COUNT 476 a 601 c 500 g 551 t

ORIGIN

Query Match 54.6%; Score 1240.4; DB 10; Length 2128;

Best Local Similarity 84.7%; Pred. No. 0;

Matches 1433; Conservative 0; Mismatches 241; Indels 18; Gaps 3;

1 GGGGGTGAAGCATACGTTTGGCCGCAATTCGGGGCGCGGAGCTGGGGGGTCCCTGTG 60

3 GGGGTGAAGCCACCGCGTCTGCCCAATTCGGGGCGCGGAG---CTGGGGTCCCGGG 59

61 GGGCTCCGGAGTAAAGATGGGCTCTCAGCGGAGGGGAGCGAGGGAGCTGGTGGCGC 120

60 GGGGGCCGCAATTAAGATGGGCTCTCAGCGGAGCGGAGCGGAGAGAGAGCGCTGGAGC 119

121 TGGCGGGGTTTCGAGTCGGGTTTCAGAGCTGAGCTTAAACAAGTTGGGAGCTGCC 180

120 AGATCGGGGGCTGGCGGGGTGTCAGGATCTGAGCTTAAACAAGTTGGGAGCTGTC 179

181 TGGGGCGCTCAGAACAGCGCTGCGGCTGATCTTCATCTTCTGGGTTAACCCCTTGG 240

180 TGGGGCGCTCAGAACAGCGCTGCGGCTGATCTTCATCTTCTGGGCTAACCCGTTGG 239

241 CTGTGTTTATGGGATTACCTTTCTAACAGAGACCTACTCATCCACCTCTTCACATA 300

240 CTGTGTTTATCGGGATTACCTTTCTAACAGAGACCTACTCATCCATCTCTTCACATA 299

301 CCTTTACAGGCTCTCAATTTGTTTAACTTTTGAACAGAGCTTACACACTGCTGCG 360

300 CCTTACAGGCTCTCAATTTGTTTAACTTTTGAACAGAGCTTACACACTGCTGCG 359

361 TGTGATGTTGCTTCAAGTTCCTTCACTTCTGACTAATGGGCGGACACATGCTGCGTTC 420

360 TATGTTGTTGCTTCAAGTTCCTTCACTTCTGACTAATGGGCGGACACCTGCTGCGTTC 419

421 TCACATACCTTTGCTTCCAGATGGGCTACTCTGGCTGGATACATTAACAATGACACCG 480

420 TTACATACCTTTGCTTCCAGATGGGCTACTCTTGGCGGATATTAACAACAAGACACCG 479

481 GCAACTAGCATATCAAGTGAACATGCCAATTTGTTTCTGACTTTGAAGCTGATGTT 540

480 GTGACTAGCATATCAAGTGAACATGCCAATTTGTTTCTGACTGAACATGAAGCTGATGTT 539

541 TGGCTGTTGACTTACTTTGACGAGGAGAAAGATCAAAATTCCTTGTCTCTGACACAGA 600

540 TGTGATGTTGACTTACTTACATGAGGACCAAGAGCGGAAATTCCTTGTGACTCTGACACAGA 599

601 AATATGCATACGTGCTGCTTCCCTGCGAGAGTGGCTGCTTCTCTACTTCTATG 660

600 AATATGCATACGTGCTGCTTCCCTGCGAGAGTGGCTGCTTCTCTACTTCTATG 659

661 GGGCTTCTTGTGAGGGCCCGAGTTCTCAATGAATCACTACATGAGCTGTGACGAG 720

660 GAGCTTCTTGTGAGGGCCCGAGTTCTCAATGAATCACTACATGAGCTGTGACGAG 719

721 AGCTGATTGACATACAGAAAGATACCAACAGCATATTCCTGCTCAAGCGCTGTA 780

720 AGCTGATTGACATACAGAAAGATACCAACAGCATATTCCTGCTCAAGCGCTGTA 779

781 GCTGGGCTTCTTCTACTACTGAGGCTACACAGCTGCTGAGCCCGACATCAAGAGCT 840

780 GCTGGGCTTCTTCTACTACTGAGGCTACACAGCTGCTGAGCCCGACATCAAGAGCT 839

841 ATCTCTCACTGAGAGCTATGACAAACCAACCTTCTGCTGCTGCTGATGATGATGTA 900

840 ATCTCTCACTGAGAGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 899

901 TCTGGGCAAGTTGTGCTGATCAAAATATGTCACCTGTTGCTGCTGATCAAGAGAGTAT 960

900 TCTGGGCAAGTTGTGCTGATCAAAATATGTCACCTGTTGCTGCTGATCAAGAGAGTAT 959

961 GCATTTTGAAGGGGCTGCTTCAATGGCTTTGAAGAAAGGCAAGGCAAGGAGGATG 1020

960 GCATTTTGAAGGGGCTGCTTCAATGGCTTTGAAGAAAGGCAAGGCAAGGAGGATG 1019

1021 CCTGTGCCAATGAGAGTGTGCTCTTTGAAGAAAGGCAAGGCAAGGCAAGGATG 1080

1020 CCTGTGCCAATGAGAGTGTGCTCTTTGAAGAAAGGCAAGGCAAGGCAAGGATG 1079

1081 CCTATTCAATCAACCAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140

1080 CCTATTCAATCAACCAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1139

1141 TCTTTGAAATTAAGAACTCTCAAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200

1140 TCTTTGAAATTAAGAACTCTCAAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1199

1201 GCTTGCATCAGAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260

1200 GCTTGCATCAGAGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259

1261 GACAGCTGCAAGCTCAATTAAGAGAGCCCACTGAGCAAGCTGAGCAAGCTGAGCA 1320

1260 GACAGCTGCAAGCTCAATTAAGAGAGCCCACTGAGCAAGCTGAGCAAGCTGAGCA 1319

1321 TCTTCCAGCCCTTCTACTATTTGAGCAACAGACATCACTGCTGCTTCAATGGTTACT 1380

1320 TCTTCCAGCCCTTCTACTATTTGAGCAACAGACATCACTGCTGCTTCAATGGTTACT 1379

1381 CCATGACTGCTTCTGCTTCTTCAAGTGGAGCAATAGGTTAAGTGTATTAATCATCT 1440

1380 CCATGACTGCTTCTGCTTCTTCAAGTGGAGCAATAGGTTAAGTGTATTAATCATCT 1439

1441 ATTTCCTTGGCAATCTTCTTCTGAGCTTCAATTAATGCTTATTAATTAATTAAT 1500

1440 ATTTCCTTGGCAATCTTCTTCTGAGCTTCAATTAATGCTTATTAATTAATTAAT 1499

1501 CAATGAGTCCAAAGAAAGAGATTAAGAAAGAGATTAATC-----CATTTCCCTG 1555

1500 CAATGAGTCCAAAGAAAGAGATTAAGAAAGAGATTAATC-----CATTTCCCTG 1554

1556 GAGCTGTGCGGAGCTGTGAGAAATCTAGCTGCTTCTTCAAGAGCTCTTGTGCGC 1615

1555 GAGCTGTGCGGAGCTGTGAGAAATCTAGCTGCTTCTTCAAGAGCTCTTGTGCGC 1614

1616 CAGAGCAAGATGAGAAAGCCAGGAGAGTGAAGATGATGATGATGATGATGATGATG 1675

1615 CAGAGCAAGATGAGAAAGCCAGGAGAGTGAAGATGATGATGATGATGATGATGATG 1674

1620 -----AGCAGAGAGCCAGAGAGTGAAGATGATGATGATGATGATGATGATG 1669

QY 1676 CTCGACGCCAAG 1687
Db 1670 CGACGCCAAG 1681

RESULT 11
BC006753

DEFINITION
BC006753 1898 bp mRNA linear ROD 16-APR-2003
MGC:11670 IMAGE:3709076, complete cds.

ACCESSION
BC006753

VERSION
BC006753.1 GI:13879545

KEYWORDS
MGC.

SOURCE
Mus musculus (house mouse)

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 1898)
Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Adams, R.D., Mullan, S.J., Bosak, S.A., McSwan, P.J.,
McKernan, K.J., Malek, J.A., Guarnatone, P.H., Richards, S.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahy, J., Helton, E., Kettelman, M., Madan, A., Young, A.C., Shvachenko, Y.,
Sanchez, A., Whiting, M., Madan, A., Madan, A.C., Rodriguez, S.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E.,
Schmeck, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE
22388257

JOURNAL
MEDLINE
PUBMED
12477932
2 (bases 1 to 1898)

REFERENCE
Strausberg, R.
Direct Submission
Submitted (27-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
USA

JOURNAL
COMMENT
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Guarnatone, P.H., Garcia, A.M., Lu, X., Huiyik, S.W., Louisedge, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Naranvali,
A.N., Gibbs, R.A.

FEATURES
SOURCE
1..1898
/organism="Mus musculus"
/mol_type="mRNA"

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAC Plate: 17 Row: h Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarly but not
identity to protein.
Location/Qualifiers

gene
1..1898
/gene="Grc3f"
/note="synonyms: C3f, PTC"
/db_xref="LocustID:14792"
/db_xref="MGI:1315211"
34..1497
/codon_start=1
/product="gene rich cluster, C3f gene"
/protein_id="AAH06753.2"
/db_xref="GI:14548370"
/db_xref="LocustID:14792"
/translation="MASTADGDMGTLEBMRGLMPGVEDLSXKLATSLGASEOALRL
LTSIFUGFLALFYHYLPYKDSYIHLFHTPTGISATFNFGHOFHSLLCVLOFL
ILRLGRTVAVITTLCEQAYLLAGYTTAGDVDIKWTMPHCVTLKILGLCTIDY
DEGKXGNSLTSQKXAIKGVPSLLEVAGFSYFAFLVPOFSNMHWKLVKVGQGLD
IGKPNSTIPALKEISLGLVYLVGYTLISPHITDDYLTEDYDNRPFRCMGLMI
GKFLVLYKVTCLVTEGVCLISGLGNGDEGTVRMDACAMKWLFFETPRNGTI
ASFNINMAMVARYI KPLKLGKNEISGLSLFLALMGLHSGYLICFOMEFLIVI
YKQVSSILRDSPLASSLSTALDPLFYLUOTTIHMLFMGYSMTAFCLPTMDKMLV
YSILFLGVHFLSLFLIPYHKKMPVKELKRE"

CDS
BASE COUNT 420 a 538 c 448 g 492 t

ORIGIN

Query Match 54.3%; Score 1233.4; DB 10; Length 1698;
Best Local Similarity 85.3%; Pred. No. 0;
Matches 1404; Conservative 0; Mismatches 226; Indels 15; Gaps 2;

QY 48 GGGGGTCCCTGTTGGGGCTCCCGAGTTAAGATGCGTCTCAGCGGAGGGAGCGAGG 107
Db 4 GGGGGTCCCGGGGGGGGGCGCGAGTTAAGATGCGTCTCAGCGGAGCGGAGCATGGGA 63

QY 108 ACTGTGTGTCGCTGGCGGGGGTTCGACGTGGGGTTCCAGAGCTGAGCTTAAACAAG 167
Db 64 GAGACCTGTGAGCAGATGCGGGGGCTGTGGCGGGGTTCGAGGATCTGAGCTTAAACAAG 123

QY 168 TTGGGAGCTCCCTGGGGGGCTGAGAACAGGGCTGGGGCTATCATCTCCATCTTCGG 227
Db 124 TTGGCAGCTCTTGAGCGGTGCGAACAGGGCTGGGCTCATCTTCTCATCTTCG 183

QY 228 GGTACCCCTTGTGCTTTTATCCGACATTAACCTTTCTACAGAGAGACTTACCTTCATC 287
Db 184 GGTACCCCTTGTGCTTTTATCCGACATTAACCTTTCTACAGAGAGACTTACCTTCATC 243

QY 288 CACCTTTTCATACCTTTACAGGCTCTCATATGCTTATTTTACCTTTGAAACGAGCTC 347
Db 244 CATCTTTCACACCTTTCAGGGCTCTCATATGCTTATTTTACCTTTGAAACGAGCTC 303

QY 348 TACCACTCCCTGTGTGATGTCCTGACAGTTCCTCATCTTCGATTAATGGGCGGACG 407
Db 304 TACCACTCCCTGTGTGATGTCCTGACAGTTCCTCATCTTCGATTAATGGGCGGACG 363

QY 408 ATCACTGCGTCTCTCACTACCTTTTCTTCAGATGAGCTTACCTTGTGCTGATATAT 467
Db 364 GTCACTGCGTATTAATCACTACCTTTCTTCAGATGAGCTTACCTTGTGCGGATATTA 423

QY 468 TACACTGCAACGGCACTACATCAAGTATCAAGTGAATGACATGCTGTGCTGACTTGG 527
Db 424 TACACAGCAACCGGTACTACATCAAGTGAATGACATGCTGTGCTGACTTGG 483

QY 528 AACCTGATGTGCTGTGCTGCTACTTGAAGGAGGGAAGATACAGATTCCTGTGTC 587
Db 484 AACCTGATGTGCTGTGCTGCTACTACAGAGGAGGGAAGATTCCTGTGTC 543

QY 588 TCTGAGCAACAGAAATATGACATACGTGTGCTTCTTCCCTGTGGAAGTGTGCTGCTTC 647

Db 544 TCTGACCAACAGAAATATGCCATACGGGGTGTCCCTTCATGTGCGAGTTGCTGCTTC 603
 Qy 648 TCTTACTTATAGGGGCTTCTGTAGGGGCCCACTTCTCATATGATCATCATGAAG 707
 Db 604 TCTTACTTATAGGGGCTTCTGTAGGGGCCCACTTCTCATATGATCATCATGAAG 663
 Qy 708 CTGGTGCAGGAGAGCTGATTTGACATACAGAAAGATACCAAGCATCATCTTCTGCT 767
 Db 664 CTGGTGCAGGAGAGCTGATTTGACATACAGAAAGATACCAAGCATCATCTTCTGCT 723
 Qy 768 CTCAAGCGCTGAGTGTGGGCTTCTCTACATAGTGGCTTACATCTGCTGACCCCTCC 827
 Db 724 CTCAAGCGCTGAGTGTGGGCTTCTCTACATAGTGGCTTACATCTGCTGACCCCTCC 783
 Qy 828 ATCAGAGAGACTATCTCTCACTGAGACTATGACCAACACCCCTTCTGTTCCGCTGC 887
 Db 784 ATCAGAGAGACTATCTCTCACTGAGACTATGACCAACACCCCTTCTGTTCCGCTGC 843
 Qy 888 ATGTACATCTGATCTGGGCAAGTTGTGTCTGTCATAAATATGTCACTGTGGCTGTC 947
 Db 844 ATGTACATCTGATCTGGGCAAGTTGTGTCTGTCATAAATATGTCACTGTGGCTGTC 903
 Qy 948 AAGAGAGAGTATGATTTTGAAGGGCTGCTGATGAGCTTTTGAAGAAAGGCGAAG 1007
 Db 904 AAGAGAGAGTATGATTTTGAAGGGCTGCTGATGAGCTTTTGAAGAAATGGGACC 963
 Qy 1008 GCAAGTGGGATGCTGCTGCTCAACATGAGTGTGGCTTTTGAACAAACCCCGCTTC 1067
 Db 964 GTGAGATGGGATGCTGCTGCTCAACATGAGTGTGGCTTTTGAACAAACCCCGCTTC 1023
 Qy 1068 ACTGACCATTTGCTGCTTCAACATCAACCAACGCTGGGTGGCCCGCTACATCTTC 1127
 Db 1024 AATGACCATTTGCTGCTTCAACATCAACCAACGCTGGGTGGCCCGCTACATCTTC 1083
 Qy 1128 AAGAGATCAATGCTTCTGGAATTAAGAACTCTCTGAGGCTCTCTGTTCTGATTTCTG 1187
 Db 1084 AAGAGATCAATGCTTCTGGAATTAAGAACTCTCTGAGGCTCTCTGTTCTGATTTCTG 1143
 Qy 1188 GGCCTTGGGACGGCTGCTGCTGAGATACCTGCTGCTTCCAGATGGAAATTCCTCAT 1247
 Db 1144 GGCCTTGGGACGGCTGCTGCTGAGATACCTGCTGCTTCCAGATGGAAATTCCTCAT 1203
 Qy 1248 GTTATTTGGAAGAAGAGCTGCCAGGCTCATTTCAAGAGAGGCCCACTGAGCAAGCTG 1307
 Db 1204 GTTATTTGGAAGAAGAGCTGCCAGGCTCATTTCCGGAAGAGCCCTGAGCAAGCTG 1263
 Qy 1308 GCGGCAATTAATCTCTCTCAAGCCCTTCTAATTTGGTGAACAGACATCACTGAGCTC 1367
 Db 1264 GCGGCAATTAATCTCTCTCAAGCCCTTCTAATTTGGTGAACAGACATCACTGAGCTC 1323
 Qy 1368 TTCATGGGTTACTCCATGATGCTTCTGCTTCAAGTGGGACAAATGGCTTAAGTG 1427
 Db 1324 TTCATGGGTTACTCCATGATGCTTCTGCTTCAAGTGGGACAAATGGCTTAAGTG 1383
 Qy 1428 TATTAATCATTAATTTCTTGGGCAATCTTCTGAGCTTACTATTTATTTGCT 1487
 Db 1384 TACAGATCATTAATTTCTTGGGCAATGCTTCTTCTGAGCTTACTATTTATTTGCT 1443
 Qy 1488 TATATTCACAAAGCAATGTGTCCAAAGAAAGAAAGTAAAGAGATGAATTAATC--- 1543
 Db 1444 TATATTCACAAAGCAATGTGTCCAAAGAAAGAAAGTAAAGAGATGAATTAATC--- 1503
 Qy 1544 -CATTTCCCTGGTGTGGGCTGAGGAGTGTGACAGAAATTAATCTGCTTCTTCTTCAAG 1602
 Db 1504 TCTTTTCCCTGGTGTGGGCTGAGGAGTGTGACAGAAATTAATCTGCTTCTTCTTCAAG 1563
 Qy 1603 CACTCTTTTCCCAAGAGCAAGAAATGAAAGCAAGGAGGTGGAAGATCATCTTCC 1662
 Db 1564 CACTCTCTCAAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1613
 Qy 1663 AGCTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1687
 Db 1614 AGCTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1638

RESULT 12
 AY028317
 LOCUS 1668 bp mRNA linear ROD 01-JUN-2001
 DEFINITION Mus musculus putative transmembrane protein PTG mRNA, complete cds.
 ACCESSION AY028317
 VERSION AY028317.1 GI:14278836
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
 REFERENCE
 TITLES
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source
 1. 1968
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /chromosome="6"
 /tissue_type="liver"
 /note="ppar alpha target"
 /codon_start=1
 /product="putative transmembrane protein PTG"
 /protein_id="AAK20915.1"
 /db_xref="GI:14278837"
 /translation="MASTADDMGSELTSEOMGLMPGVEDLSLNKATSLAGSEALRL
 IFSITFLGRLALFPHRYLPHYKDYSLIHLPHFTGSLSTYRPNFGHSLSLCVALQFL
 IRLMGRVLAIVITLCPQMAVLLAGVYVTAAGDDIKMTWPHCVLTLLGLCIDY
 DGGDGNLSLTSBOQKLAIRGVPSLLEVAFLVGDQFSNMHYMKLVKGLTD
 IPGKPNSTIIPALKSLGLVLYLVSHPHITDYLITEDYDNRPFMRPMYMLI
 GKFVLYKVTCLMVTGVCILSGDFNFGDSENGTIRVMDACAMKXMLEPTEPENGIT
 ASPNINAMVARYIFRKLKPLGNESLQGLSLFLALMGLHSGYLICPQEPFLIV
 VEKQVSLIRSPALSLASITTAQPIYLYVQOTTHLPMGYSMTAFCLFTWDMKLK
 YRSIFYLGHVFFLSLFTLPYIHKMVRKRELKRE"
 BASE COUNT 443 a 554 c 469 g 502 t
 ORIGIN
 Query Match 54.1%; Score 1229.4; DB 10; Length 1668;
 Best Local Similarity 84.6%; Pred. No. 0; Mismatches 241; Indels 19; Gaps 4;
 Matches 1433; Conservative 0;

1 GGGGGTGAAGCATTCGTTTGGCCCGCATTCGGGGGCGGCGGATGGGGGGTCCCTGTG 60
 Db 8 GGGGTGAAGCAACCGCTTCTGCCACATTCGGGGGCGGAG--CTGGGGGTCCCCGGG 64
 Qy 61 GGGCTCCGGAATTAAATGCGCTCTCAGCGGAGGGGAGCAGAGGGACTGTGTGGCGC 120
 Db 65 GGGCGCCGCGAGTTAATGCGCTCTCAGCGGAGGGGAGCATGGAGAGAGCGTGGAGC 124
 Qy 121 TGGCGGGGCTTTCAGAGTGGGTTTCCAGAGCTGAGCTTAAAGAGTTGGCGAGCTCC 180
 Db 125 AGATGCGGGGCTGTGGCGGGGTGTGAGGATCTGAGCTTAAAGAGTTGGCGAGCTTC 184
 Qy 181 TGGCGGCTCAGAACAGGCGCTGCGGCTGATCATCTTCCTGAGTTAACCCCTTGG 240
 Db 185 TGGGGCGCTCGGAACAGGCGCTGCGGCTCATCTTCTCATCTTCTGAGCTAACCCGTTGG 244
 Qy 241 CTTTGTTTTATGCGATTACTCTTTTCAAGAGAGACCTTCACTTCACTCTTTCATA 300
 Db 245 CTCTGTTTACCGGCTTACTTCTTCTCAAGAGAGACCTTCACTTCACTCTTTCACA 304

QY	301	CCCTTACAGGGCCTCTCAATTCCTTATTTTAACTTTGGAAACCACTCTACACCTCCCTGC	360
Db	305	CCTTACAGGGCCTCTCAATTCCTTATTTTAACTTTGGAAACCACTCTACACCTCCCTGC	364
QY	361	TGTGATATGTGCTTCAGTTTCTCTCATTCCTTGCACATATGAGGGCCGACATCATCCGCTCC	420
Db	365	TATGATGTGCTTCAGTTTCTCTCATTCCTTGCACATATGAGGGCCGACATCATCCGCTCC	424
QY	421	TCACCTACCTTTTGGCTTCCAGATGGCCCTTACCTTGGCTGGATATCTATTTACCTGGCACCG	480
Db	425	TTACTACCTTTTGGCTTCCAGATGGCCCTTACCTTGGCTGGATATCTATTTACCTGGCACCG	484
QY	481	GCAACTACGATATCAAGTGAACAATATGCCAATTGTGTTCTACCTTTGAGCTGATTTGTT	540
Db	485	GTGACTACGATATCAAGTGAACAATATGCCAATTGTGTTCTACCTTTGAGCTGATTTGTT	544
QY	541	TGGCTGTTGACTACTTTTGACGAGGAGGAAGATTCAGAAATTCCTTGTCCCTTGAGCAAGA	600
Db	545	TGTGATTTGACTACTAGATGGAAGCAAGAGCGGAATTCCTTACCTCTGAGCAAGA	604
QY	601	AATATGCCAATACGTGATGTTTCTTCCCTGCTGGAAATTGCTGTTTCTCTACTTTCTATG	660
Db	605	AATATGCCAATACGTGATGTTTCTTCCCTGCTGGAAATTGCTGTTTCTCTACTTTCTATG	664
QY	661	GGGCGCTTGTGGTGGGGCCGAGTTCTCAATGAATCATATAGATGAGCTGTGAGGGAG	720
Db	665	GAGCCTTCTGGTGGGGCCGAGTTTCAATGAATCATATAGATGAGCTGTGAGGGAG	724
QY	721	AGCTGATTTGACATACAGAGAAAGATATCCAAACAGATCATTCCTGCTCTCAAGCGCTGA	780
Db	725	AGCTGATTTGACATACAGAGAAAGATATCCAAACAGATCATTCCTGCTCTCAAGCGCTGA	784
QY	781	GTCGTGGCCTTTTCTACTAGTGGGCTTACACACTGCTCAAGCCCCCACTACAGAAAGCT	840
Db	785	GTCGTGGCCTTTTCTACTAGTGGGCTTACACACTGCTCAAGCCCCCACTACAGAAAGCT	844
QY	841	ATTCCTCAGTGAAGATATATGACAAACACCCCTCTGGTTCGGCTGATATGATCATGCTGA	900
Db	845	ATTCCTCAGTGAAGATATATGATTAACCGCCCTTCTGGTTCGGCTGATATGATCATGCTGA	904
QY	901	TCTGGGGCAAGTTTGTGCTGTATCAAAATATGTCACTGTGTGGCTGGTCAAGAAAGATAT	960
Db	905	TCTGGGGCAAGTTTGTGCTGTATCAAAATATGTCACTGTGTGGCTGGTCAAGAAAGATAT	964
QY	961	GCATTTTGAACGGGCGCTTGCTTCAATGGCTTTGAAAGAAAGGCAAGCAAGTGGGATG	1020
Db	965	GCATTTTGAACGGGCGCTTGCTTCAATGGCTTTGAAAGAAAGGCAAGCAAGTGGGATG	1024
QY	1021	CCTGTGCCAATGAAGGTGTGCTCTTTGAAACAAACCCCGCTTCACTGGACATATG	1080
Db	1025	CCTGTGCCAATGAAGGTGTGCTCTTTGAAACAAACCCCGCTTCACTGGACATATG	1084
QY	1081	CCTCATTTCAACATCAACACCAAGCGCTGGTGGCCCGCTACATCTTCAAGCATCAAGT	1140
Db	1085	CCTCATTTCAACATCAACATCAAGCGCTGGTGGCCCGCTACATCTTCAAGCATCAAGT	1144
QY	1141	TCCTTGAAGATTAAGAACTCTCTCAGGAGTCTCTCGTTGCTATCTCTGGCCCTCTGGACG	1200
Db	1145	TCCTTGAAGATTAAGAACTCTCTCAGGAGTCTCTCGTTGCTATCTCTGGCCCTCTGGACG	1204
QY	1201	GCTGTGCACTCAGAGATACCTGGTGTGCTTCCAGATGGAATTTCTCATTTTATTTGTGAAA	1260
Db	1205	GCTGTGCACTCAGAGATACCTGGTGTGCTTCCAGATGGAATTTCTCATTTTATTTGTGAAA	1264
QY	1261	GACAGAGCTGCCAGGCTCATTTAAGAGAGCCCAACCTGAGCAAGTGTGGCGGCATTTACTG	1320
Db	1265	AGCAGAGCTGCCAGGCTCATTTAAGAGAGCCCAACCTGAGCAAGTGTGGCGGCATTTACTG	1324
QY	1321	TCCTTCAGGCCCTTCTACTATTTTGTGTGCAACAGACCATTCACATGGCTCTTTCAGTGGGTTACT	1380
Db	1325	CCCTACAGGCCCTTCTACTATTTTGTGTGCAACAGACCATTCACATGGCTCTTTCAGTGGGTTACT	1384
QY	1381	CCATGACTGCTTCTGCTCTTTCACGTGGGCAAAATGGCTTATAGGTGATATAATTCATCT	1440


```

/Protein_id="AAP35646.1"
/db_xref="GI:30582839"
/tranlations="MGRITTAVALTFPCFQMAVLLAGYYTATGNDIKWMPHCVLL
KILGLADYEDGDKDONSLSBOOKXYAIGVPSLEIVAPSEYVYGAFLVGPQSNMNY
MLVQSELIDIRKIDPNSIIPALKRISLGLFLVGVTLTSPHITERYLLTREDNHPF
WRRCNMLMKRFVLYKYTCMLVTGVCILTELGNNGEERKAKWDCANMKWLF
ETNPRFTGILASININAMVARYIFKRFLGNKELSGSLFLALWHGHSGLV
CFQMKFLIVIERQARLIQSEFTLSKLAITVLQPFYLVQOTIHLFMGYSMTAF
LFTWDMKLVKYSIYFLGHIFFLSLFIPIYIKHAWVPREKRLKKE"
BASE COUNT      274 a      310 c      256 g      306 t
ORIGIN
Query Match      50.3%; Score 1143.4; DB 9; Length 1146;
Best Local Similarity 99.9%; Pred. No. 1.7e-299;
Matches 1144; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 396 ATGGGCGGACCATCTACTGCGGCTCTCACTACTCTTTTGGCTTCAGATGGCCTACCTTCG 455
DB 1 ATGGGCGGACCATCTACTGCGGCTCTCACTACTCTTTTGGCTTCAGATGGCCTACCTTCG 60
QY 456 GCTGATACCTATTACCTGACCTGCGGCACTACGATATCAAGTGCAGATGCCACATGTT 515
DB 61 GCTGATACCTATTACCTGACCTGCGGCACTACGATATCAAGTGCAGATGCCACATGTT 120
QY 516 GTTCTGACTTTGAAGCTGATTTGGCTGTGACTACTTTGACGAGGAGAAAGATCAG 575
DB 121 GTTCTGACTTTGAAGCTGATTTGGCTGTGACTACTTTGACGAGGAGAAAGATCAG 180
QY 576 AATTCCTTCTCTCTGAGCAACGAAATATGCATACGTTGCTTCTTCCCTGCTGAA 635
DB 181 AATTCCTTCTCTCTGAGCAACGAAATATGCATACGTTGCTTCTTCCCTGCTGAA 240
QY 636 GTTGTGCTTTCTCTACTTCTTATGAGGCGCTTCTTGTAGGCGGCCAGTTTCAATGAT 695
DB 241 GTTGTGCTTTCTCTACTTCTTATGAGGCGCTTCTTGTAGGCGGCCAGTTTCAATGAT 300
QY 696 CACTACATGAAGCTGTGAGGAGAGCTGATTCATACCAAGAAATACCAACACG 755
DB 301 CACTACATGAAGCTGTGAGGAGAGCTGATTCATACCAAGAAATACCAACACG 360
QY 756 ATCATTCCTGCTCTCAAGCGGCTGAGTGGGCTTTTCTACTAGTGGGCTTACACATG 815
DB 361 ATCATTCCTGCTCTCAAGCGGCTGAGTGGGCTTTTCTACTAGTGGGCTTACACATG 420
QY 816 CTCAGCGCCCACTACAGAAAGCTATCTCTCACTGAAAGACTATGACCAACCCCTTC 875
DB 421 CTCAGCGCCCACTACAGAAAGCTATCTCTCACTGAAAGACTATGACCAACCCCTTC 480
QY 876 TGGTTCGCTGATGATGCTGATCTGGGCAAGTTTGTGCTGACAAATATGTCACC 935
DB 481 TGGTTCGCTGATGATGCTGATCTGGGCAAGTTTGTGCTGACAAATATGTCACC 540
QY 936 TGTGGCTGTGACAGAAAGATGATGATTTGACCGGCGCTTGGGCTTCAATGGCTTTGAA 995
DB 541 TGTGGCTGTGACAGAAAGATGATGATTTGACCGGCGCTTGGGCTTCAATGGCTTTGAA 600
QY 996 GAAAAGGAGGAGGAGGAGGAGTGTGCTGACCAATGAAGGTGGTCTTTGAA 1055
DB 601 GAAAAGGAGGAGGAGGAGGAGTGTGCTGACCAATGAAGGTGGTCTTTGAA 660
QY 1056 AACCCCGCTTCACTGACCAATGCTCATTTCAACATCAACCAACCCCTGGTGCC 1115
DB 661 AACCCCGCTTCACTGACCAATGCTCATTTCAACATCAACCAACCCCTGGTGCC 720
QY 1116 CGCTCATTTTCAAGAGACTCAAGTTCTTTGAAATTAAGAACTCTCTAGGGTCTCTCG 1175
DB 721 CGCTCATTTTCAAGAGACTCAAGTTCTTTGAAATTAAGAACTCTCTAGGGTCTCTCG 780
QY 1176 TTGCTATTCCTGGCCCTTGACACGGCTGACACTGAGATACCTGGTCTGTTCCAGATG 1235
DB 781 TTGCTATTCCTGGCCCTTGACACGGCTGACACTGAGATACCTGGTCTGTTCCAGATG 840
QY 1236 GAATTCCTCATTTGTTATTTGTGAAAGACAGGCTGCCAGGCTCATTTCAAGAGGCCAC 1295

```

```

DB 841 AATTCCTCATTTGTTATTTGTGAAAGACAGGCTGCCAGGCTCATTTCAAGAGGCCAC 900
QY 1296 CTGAGAACCTGCGCCCATTTACTGCTCTCCAGGCTTTCTACTATTTGGTCAACAGACC 1355
DB 901 CTGAGAACCTGCGCCCATTTACTGCTCTCCAGGCTTTCTACTATTTGGTCAACAGACC 960
QY 1356 ATCCACTGCTCTTCATGAGGTACTGCTCAGATGCTTGGCTCTTCAGTGGGACAA 1415
DB 961 ATCCACTGCTCTTCATGAGGTACTGCTCAGATGCTTGGCTCTTCAGTGGGACAA 1020
QY 1416 TGGCTTAAGGTATTAATATCCATCTATTCTTGGCCACATCTTCTCTGAGCCTACTA 1475
DB 1021 TGGCTTAAGGTATTAATATCCATCTATTCTTGGCCACATCTTCTCTGAGCCTACTA 1080
QY 1476 TTCAATTTGCTTTATTTATTCACAAAGCATGTGTGCCAAGAAAGAGATTAAAGATG 1535
DB 1081 TTCAATTTGCTTTATTTATTCACAAAGCATGTGTGCCAAGAAAGAGATTAAAGATG 1140
QY 1536 GAATA 1540
DB 1141 GAATA 1145
RESULT 14
BT007735 1146 bp mRNA linear SYN 13-MAY-2003
LOCUS BT007735
DEFINITION Synthetic construct Homo sapiens putative protein similar to nesby
(Drosophila) mRNA, partial cds.
ACCESSION BT007735
VERSION BT007735.1 GI:30584308
KEYWORDS FLI CDNA.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 1146)
AUTHORS Kauline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Kauline,N., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y.,
Pheilan,M. and Farmer,A.
TITLE Cloning of human full-length CDSs in BD Creator(TM) System Donor
vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1146)
AUTHORS Kauline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Kauline,N., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y.,
Pheilan,M. and Farmer,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
Circle, Palo Alto, CA 94303, USA
COMMENT This CDS clone is a part of a collection of human full length
expression clones generated by BD Biosciences Clontech and the
Harvard Institute of Proteomics. Each CDS has been cloned in two
forms: with and without stop-codon (to allow fusion with C-terminal
tag). The CDS has been directionally cloned using BD in-Fusion(TM)
cloning system between the SalI and HindIII sites of the pDNR-DUAL
vector. Additional sequences in the clone: 'ACC' after SalI site
and before 'ATG' to provide kozak consensus sequence; 'GG' after
last codon and before HindIII site to maintain reading frame.
Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES
source
1..1146
location/Qualifiers
1..1146
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="GH00356L1.0"
/clone_lib="BD Creator(TM) CDS Library derived from MGC
collection"
/lab_host="DH5alpha T1 resistant"
/note="vector: pDNR-Dual"
1..>1146
/note="Mutations: 1145:Stop->Leu"
/codon_start=1
/transl_table=11

```

/product="Homo sapiens putative protein similar to mesy
(Drosophila)"
/protein_id="AA036403.1"
/db_xref="GI:30584309"
/translation="MGRTTAVLTTFQFQVAVLLAGYYTATGNVDIKMTMPCVLT
KLGLAVDFEDGKDNLSSEQKYARIVPSSLLEAGASYGAFVQPSNMHY
MKLVQELIDIPKIPNSIIPALRSLISFLYVLTSLPHTEDYLIEDNDHF
WRCQWMLIWGKVLKYVTCMLVTEGVCITLGLFNGPEKPKAKMDACAMKWLF
ENMPRTGTIASPININMNAVARYIPRRLKFLNKEISQSLPLALNHLGSLV
CFQMKFLIVERQAPARLQIESPFLSLAIVLPPIYLVQDTIMHLENGISMTAF
LFTWDMKLVKYSIYFLGHIPLFLFLIYPIHKAVPRREKUKRMEL"

BASE COUNT 273 a 310 c 256 g 307 t

ORIGIN

Query Match 50.3%; Score 1142.4; DB 12; Length 1146;
Best Local Similarity 99.98; Pred. No. 3.2e-299;
Matches 1143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

396 ATGGGCGGACACATCACTGCGGCTCTCACTACCTTTGGTTCAGATGGCTACCTCTG 455
1 ATGGGCGGACACATCACTGCGGCTCTCACTACCTTTGGTTCAGATGGCTACCTCTG 60
456 GCTGATATTATTACATCTGCCACCGGCACTACGATATCAAGTGAACATGCGCATTTG 515
61 GCTGATATTATTACATCTGCCACCGGCACTACGATATCAAGTGAACATGCGCATTTG 120
516 GTTCTGACTTTGAAGCTGATTTGGTGGCTGTGACTACTTTGAACGAGGAAAGATCAG 575
121 GTTCTGACTTTGAAGCTGATTTGGTGGCTGTGACTACTTTGAACGAGGAAAGATCAG 180
576 AATTCCCTGTCTCTGAGCAACAGAAATATGCCATACGTGTGCTCTCCCTGCTGAGAA 635
181 AATTCCCTGTCTCTGAGCAACAGAAATATGCCATACGTGTGCTCTCCCTGCTGAGAA 240
636 GTTCTGACTTTCTCTCACTTTCTATGGGCGCTTCTGTGAGGCGCCCACTTCTCAATGAT 695
241 GTTCTGACTTTCTCTCACTTTCTATGGGCGCTTCTGTGAGGCGCCCACTTCTCAATGAT 300
696 CACTTACATGAAGCTGTGAGGAGGAGAGCTATGACATACAGAGAAAGATACCAACAGC 755
301 CACTTACATGAAGCTGTGAGGAGGAGAGCTATGACATACAGAGAAAGATACCAACAGC 360
756 ATCATTCCTGCTCTCAAGCGCCTGAGTGTGGGCTTTTCTACCTAGTGGGCTACACACTG 815
361 ATCATTCCTGCTCTCAAGCGCCTGAGTGTGGGCTTTTCTACCTAGTGGGCTACACACTG 420
816 CTCAGCCCCACATCAAGAAAGCTATCTCTCACTGAGAGCTATGACCAACACCCCTTC 875
421 CTCAGCCCCACATCAAGAAAGCTATCTCTCACTGAGAGCTATGACCAACACCCCTTC 480
876 TGGTTCGCGTCATGATCATGCTGATGTGGGCAAGTTTGTCTGTAACAAATATGTCACC 935
481 TGGTTCGCGTCATGATCATGCTGATGTGGGCAAGTTTGTCTGTAACAAATATGTCACC 540
936 TGTGCTGTGTCACAGAGAGATGATGATTTTGAACGCGGCTCAATGGCTTTGAA 995
541 TGTGCTGTGTCACAGAGAGATGATGATTTTGAACGCGGCTCAATGGCTTTGAA 600
996 GAAAGGGGCAAGGCAAGTGGATGCTGTGCAACATGAAGTGTGGCTTTTGAACA 1055
601 GAAAGGGGCAAGGCAAGTGGATGCTGTGCAACATGAAGTGTGGCTTTTGAACA 660
1056 AACCCCGCTTCTACAGGACATTTGCTCATTTCAACATCAACACCAAGCCTGTGGTGGCC 1115
661 AACCCCGCTTCTACAGGACATTTGCTCATTTCAACATCAACACCAAGCCTGTGGTGGCC 720
1116 CGCTACATCTTCAAGACACTCAAGTCTTGAAGAAATTAAGAACTCTCAAGGCTCTCG 1175
721 CGCTACATCTTCAAGACACTCAAGTCTTGAAGAAATTAAGAACTCTCAAGGCTCTCG 780
1176 TTGCTATTCCTGGCCCTCTGGAACGCGCTGACCTCAGAGATACCTGTCTGCTTCAAGATG 1235
781 TTGCTATTCCTGGCCCTCTGGAACGCGCTGACCTCAGAGATACCTGTCTGCTTCAAGATG 840

QY 1236 GAATTCCTCATTTGTTATTTGTGAAAGACAGAGCTGCCAGGCTCATTTCAAGAGACCCACC 1295
DB 841 AATTTCTCATTTGTTATTTGTGAAAGACAGAGCTGCCAGGCTCATTTCAAGAGACCCACC 900
QY 1296 CTGAGCAAGCTGGCCGCCATTAATCTGCTCTCCAGCCCTTCTACTATTGTTGCAACAGACC 1355
DB 901 CTGAGCAAGCTGGCCGCCATTAATCTGCTCTCCAGCCCTTCTACTATTGTTGCAACAGACC 960
QY 1356 ATCCACTGCTCTTCAATGGGTTACTCATGACCTGCTTCTGCTCTTCACTGGAACAA 1415
DB 961 ATCCACTGCTCTTCAATGGGTTACTCATGACCTGCTTCTGCTCTTCACTGGAACAA 1020
QY 1416 TGGCTTAAGGTGATTAATCAATCTATTTCTTGGCCACATCTTCTCTGAGCCCTACAT 1475
DB 1021 TGGCTTAAGGTGATTAATCAATCTATTTCTTGGCCACATCTTCTCTGAGCCCTACAT 1080
QY 1476 TTGATATTCCTTATATTCACAAAGCATGTGTCACAAAGAGAGATTAAGAAGATG 1535
DB 1081 TTGATATTCCTTATATTCACAAAGCATGTGTCACAAAGAGAGATTAAGAAGATG 1140
QY 1536 GAAT 1539
DB 1141 GAAT 1144

RESULT 15
AC006512/c
LOCUS
DEFINITION
Homo sapiens 12 PAC RP3-461P17 (Rowell Park Cancer Institute Human PAC library) complete sequence.
VERSION
AC006512
AC006512.13 GI:29469488
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 155975) Bp DNA linear PRI 04-APR-2003
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T., Barbarella,J., Benton,J., Bimaye,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Burch,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chin,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Hawla,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulik,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L.J., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Pay,L.L., Quiles,M., Ren,Y., Peters,L., Picken,R., Primus,E., Puy,L.L., Quiles,M., Ren,Y., Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shooshitari,N., Sisson,I.,

Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, J., Svaltek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Taney, J., Taylor, C., Taylor, T., Telirod, B., Thomas, N., Thomas, S.,
 Usami, K., Vaequez, L., Vera, V., Valladon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Kuchelapatt, R.,
 Weinstock, G. and Gibbs, R.

TITLE
 Direct Submission
 JOURNAL
 Unpublished
 REFERENCE
 2 (bases 1 to 155975)
 AUTHORS
 Worley, K.C.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (05-FEB-1999) Molecular and Human Genetics, Baylor
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 155975)
 REFERENCE
 Worley, K.C.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (29-MAY-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 4 (bases 1 to 155975)
 REFERENCE
 Worley, K.C.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (02-JUN-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 5 (bases 1 to 155975)
 REFERENCE
 Worley, K.C.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (03-JUN-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 6 (bases 1 to 155975)
 REFERENCE
 Worley, K.C.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (18-JUN-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 7 (bases 1 to 155975)
 REFERENCE
 Worley, K.C.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (27-OCT-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 8 (bases 1 to 155975)
 REFERENCE
 Worley, K.C.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (29-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 9 (bases 1 to 155975)
 REFERENCE
 Worley, K.C.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (02-APR-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 10 (bases 1 to 155975)
 REFERENCE
 Worley, K.C.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (04-APR-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Apr 2, 2003 this sequence version replaced gi:4926863.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

COMMENT
 CLONE LENGTH: This sequence does not necessarily represent the
 entire insert of this clone. Overlapping regions of clones are only
 sequenced and submitted once, so the sequence for the remainder of
 the insert may be found in the record for the adjacent clones.
 Overlapping clones are noted at the beginning and end of the
 Features listing.

ANNOTATION OF FEATURES:
 STS are identified using ePCR (genome Res. 7:541-550) searches
 of a local database that includes entries from dbSTS, GDB, and
 local mapping efforts.
 Repeats are identified using RepeatMasker (A. Smit and P. Green,
 unpublished) for Human and Mouse sequences.
 Genes and Region of sequence similarity are identified by BLAST
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
 EST and cDNA sequences. Genes demonstrate at least two exons
 flanked by consensus splice sites that maintained sequence
 continuity across the splice junctions. Sequences that are not
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
 standard of double strand coverage with a minimum of 2 clones and 2
 reads with no ambiguities or 2 chemistries with a minimum of 2
 clones and 3 reads with no ambiguities. If the sequence quality for
 a region does not meet this standard, it will be indicated in the
 annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
 standards - estimated error rate less than 1 per 10,000 bases.
 Reports of lowest quality individual bases and measures of base
 quality are listed below. Description of the metrics can be found
 at URL:
<http://www.hgsc.bcm.tmc.edu:8086/quality.info/genbank.annotation.ht>
 ml.

FEATURES

source

Location/Qualifiers

1..155975
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="12"
 /clone="RP3-461P17"
 1..97210
 /note="overlaps bases 127585..222930 of clone"
 /function="clone overlap"

misc_feature

repeat_region
 /rpt_family="13b"
 complement(2781..3065)
 /rpt_family="AluYd2"

repeat_region
 3127..3430
 /rpt_family="AluX"

STS
 3473..4048
 /standard_name="GDB:214837"

repeat_region
 3592..3797
 /rpt_family="MIR"

repeat_region
 4550..4771
 /rpt_family="13b"

repeat_region
 complement(4816..5127)
 /rpt_family="AluSp"

repeat_region
 complement(5146..5288)
 /rpt_family="FLM_C"

repeat_region
 complement(5322..5506)
 /rpt_family="AluX"

repeat_region
 5541..5833
 /rpt_family="Aluub"

STS
 6770..6912
 /standard_name="GDB:384728"

STS
 7448..7601
 /standard_name="D12S1899"

STS
 7675..7889
 /standard_name="G10506"

repeat_region
 7717..7746
 /rpt_family="(TTTA)n"

STS
 7769..7860
 /standard_name="SHGC-58828"

repeat_region
 9529..9570
 /rpt_family="(TG)n"

repeat_region
 complement(10903..11196)
 /rpt_family="AluX"

repeat_region
 12008..12158

```
repeat_region      /rpt_family="CT-rich"  
12166..12206  
repeat_region      /rpt_family="(TCTCCC)n"  
12368..12585  
repeat_region      /rpt_family="(CGGG)n"  
12647..12685  
repeat_region      /rpt_family="GC-rich"  
12822..12850  
repeat_region      /rpt_family="GC-rich"
```

```
Query Match      31.2%; Score 708.6; DB 9; Length 155975;  
Best Local Similarity 99.4%; Pred. No. 8.7e-181;  
Matches 711; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1546 TTTCCTGTGTGCTGTGTGCGGAGCTGTGTGAGAACTACTGTCTCTCTTTTCACAGCAG 1605  
DB 61329 TCTTGACAGGTGCTGTGTGCGGAGCTGTGTGAGAACTACTGTCTCTCTTTTCACAGCAG 61270  
  
QY 1606 TCCTTTGCCCCAGAGAGAGAAATGAAAAAGCCAGGAGGTGAAATCGATGCTTCCAGC 1665  
DB 61269 TCTTTTGCCCCAGAGAGAGAAATGAAAAAGCCAGGAGGTGAAATCGATGCTTCCAGC 61210  
  
QY 1666 TGTGCTCTGTCTGCTCCAGCCAGTCTTCAATTTGGGGCCAAAGGGGAAAATTTTGGAG 1725  
DB 61209 TGTGCTCTGTCTGCTCCAGCCAGTCTTCAATTTGGGGCCAAAGGGGAAAATTTTGGAG 61150  
  
QY 1726 AAGGCGTCTTGTCTTGTACCCACGCTGGAATGCAATGCGGGATCTCAGCTCACCGCA 1785  
DB 61149 AAGGCGTCTTGTCTTGTACCCACGCTGGAATGCAATGCGGGATCTCAGCTCACCGCA 61090  
  
QY 1786 CCTCCACCTCCCTGGGTTCAGATTTTCTGTGCTCAGCCTCCCAAGTAGCTGGAAATAC 1845  
DB 61089 CCTCCACCTCCCTGGGTTCAGATTTTCTGTGCTCAGCCTCCCAAGTAGCTGGAAATAC 61030  
  
QY 1846 AGGACAGCCACCAATGCCAGCTAATTTTGTATTTTCAGTAGAAACGGGATTTACACAG 1905  
DB 61029 AGGACAGCCACCAATGCCAGCTAATTTTGTATTTTCAGTAGAAACGGGATTTACACAG 60970  
  
QY 1906 TTGGCCAGGCTGTGTCTGAACTCTGTACCGCAAGTATCCACCGCTCCGCTCCCA 1965  
DB 60969 TTGGCCAGGCTGTGTCTGAACTCTGTACCGCAAGTATCCACCGCTCCGCTCCCA 60910  
  
QY 1966 GTGCTGGAGTTAAGGGGTGAGCAACCGTCCCGGCCCAAGGGGAAATCTGTGGAG 2025  
DB 60909 GTGCTGGAGTTAAGGGGTGAGCAACCGTCCCGGCCCAAGGGGAAATCTGTGGAG 60850  
  
QY 2026 GAGCAGAGGGGCTCACATCTCCCTGTGATTTCCCATGCAATGCTTATCTTCCG 2085  
DB 60849 GAGCAGAGGGGCTCACATCTCCCTGTGATTTCCCATGCAATGCTTATCTTCCG 60790  
  
QY 2086 ATCTAGCCAGGAATCTATGTGTTTTCTGTGCAATTTACTATGATGTGTATGTGCC 2145  
DB 60789 ATCTAGCCAGGAATCTATGTGTTTTCTGTGCAATTTACTATGATGTGTATGTGCC 60730  
  
QY 2146 GCTACCAACCAACCCCGCATGGGGGGGTGAGAGGGGGTGAAGGCGCTGTGCTCACT 2205  
DB 60729 GCTACCAACCAACCCCGCATGGGGGGGTGAGAGGGGGTGAAGGCGCTGTGCTCACT 60670  
  
QY 2206 TTTTCTACTTGAAGTGTATTAGATAAATCACTTGTGTTGTCAAGTTTCA 2260  
DB 60669 TTTTCTACTTGAAGTGTATTAGATAAATCACTTGTGTTGTCAAGTTTCA 60615
```

Search completed: November 22, 2003, 05:56:24
Job time : 8298 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2003, 02:05:29 / Search time 601 Seconds
(without alignments)
10209.355 Million cell updates/sec

Title: US-09-938-803-26

Perfect score: 2273

Sequence: 1 599595tgaagcgatagctt.....tttccaaaaaaaaaaaaa 2273

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_19Jun03.*
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2273	100.0	2273	22	AAC88075 Human FLEXHT-6 nuc
2	2241.2	98.6	2264	21	AAF25169 Nucleotide sequenc
3	2215	97.4	2234	21	AAZ50889 Human receptor-ass
4	1842	81.0	1842	24	ABK84420 Human cDNA differe
5	1842	81.0	1842	24	ABN96918 Gene #3416 used to
6	1842	81.0	1842	24	ABL64098 Breast cancer rela
7	1697.8	74.7	1667	23	AA570385 DNA encoding novel
8	1461	64.3	1461	22	AAF25159 Nucleotide sequenc

9	773.6	34.0	1296	23	ABV30301	Human prostate exp
10	708.6	31.2	7461	22	AAS30639	DNA encoding novel
11	708.6	31.2	7461	22	AAS28701	Genomic sequence #
12	708.6	31.2	7461	25	ACA03402	DNA encoding human
13	708.6	31.2	32174	22	ABAI1665	Human nervous syst
14	708.6	31.2	32174	22	ABAI19477	Human nervous syst
15	708.6	31.2	32174	22	ABA20359	Human nervous syst
16	708.6	31.2	32174	22	ABA21505	Human nervous syst
17	708.6	31.2	32174	22	ABJ36280	Human maculosa/kele
18	708.6	31.2	32174	22	AAS32655	Human genomic DNA
19	708.6	31.2	32174	22	AAS34394	Human DNA for a no
20	708.6	31.2	32174	22	AAS30638	DNA encoding novel
21	708.6	31.2	32174	22	AAI03792	Human reproductive
22	708.6	31.2	32174	22	AAI07447	Human reproductive
23	708.6	31.2	32174	22	AAI07491	Human reproductive
24	708.6	31.2	32174	22	AAS28700	Genomic sequence #
25	708.6	31.2	32174	22	AAI62606	Human breast or ov
26	708.6	31.2	32174	22	AAI62904	Human genomic DNA
27	708.6	31.2	32174	23	ABK72131	Human ovarian anti
28	708.6	31.2	32174	24	ABK91723	Novel ovarian anti
29	708.6	31.2	32174	25	ACA03401	DNA encoding human
30	708.6	31.2	32174	25	ABX5268	CDNA encoding nove
31	708.6	31.2	38771	22	AAK66361	Human immune/haema
32	708.6	31.2	38771	22	AAK68883	Human immune/haema
33	708.6	31.2	38771	22	AAK69604	Human immune/haema
34	708.6	31.2	38771	22	AAK71499	Human immune/haema
35	708.6	31.2	38771	22	AAK72925	Human immune/haema
36	708.6	31.2	38771	22	AAK75870	Human immune/haema
37	708.6	31.2	38771	22	AAK80325	Human immune/haema
38	708.6	31.2	38771	22	AAK83538	Human immune/haema
39	708.6	31.2	38771	25	ABZ74036	Secreted protein g
40	708.6	31.2	38771	25	ABZ74370	Secreted protein g
41	708.6	31.2	38771	25	ABZ74376	Secreted protein g
42	708.6	31.2	38771	25	ABZ67623	Human secreted pro
43	708.6	31.2	38771	25	ABZ67941	Human secreted pro
44	708.6	31.2	38771	25	ABZ67947	Human secreted pro
45	708.6	31.2	222930	24	ABK84349	Human cDNA differe

ALIGNMENTS

RESULT 1	
ID AAC88075	standard, cDNA, 2273 BP.
AC AAC88075;	
XX	
DT 09-MAR-2001	(first entry)
XX	
DE Human FLEXHT-6 nucleotide sequence SEQ ID NO:61.	
XX	
Human, FLEXHT, full-length molecules expressed in human tissue;	
KW diagnosis; gene expression; genetic linkage; genetic variability;	
KW antianemic; anticoagulant; antiarteriosclerotic; immunomodulatory;	
KW cytosolic; antisthmatic; antiinflammatory; hepatotropic; antiabiotic;	
KW anti-gout; antithyroid; neuroprotective; antiarthritic; osteopathic;	
KW epileptic; arteriosclerotic; antitumor; gene therapy; anaemia; gout;	
KW cancer; immunological disorder; atherosclerosis; developmental disorder;	
KW Crohn's disease; diabetes mellitus; bronchitis; cirrhosis;	
KW osteoarthritis; pancreatitis; rheumatoid arthritis; psoriasis;	
KW ulcerative colitis; ss.	
XX	
OS Homo sapiens.	
XX	
PN WO200070047-A2.	
XX	
PD 23-NOV-2000.	
XX	
PF 12-MAY-2000; 2000WO-US13299.	
XX	
PR 14-MAY-1999; 99US-0311894.	

PR 14-MAY-1999; 99US-0311937.
 PR 14-MAY-1999; 99US-0311940.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 PI Yue H, Tang YF, Lal P, Reddy R, Barira S, Baughn MR, Yang J;
 PI Azimtai Y, Lu DM, Au-Young U, Shih LL;
 DR MPI; 2001-016234/02.
 XX P-PSDB; AAB36584.
 PT Human FLEXHT protein and DNA sequences, useful for treating
 PT immunological disorders, developmental disorders, and cancers -
 XX
 PS Claim 5; Page 140; 168pp; English.
 XX AAC88070 to AAC88124 encode the 55 FLEXHT (full-length molecules
 CC expressed in human tissues) proteins given in AAB36579 to AAB36633. The
 CC present invention describes an isolated polypeptide (A) comprising an
 CC amino acid sequence selected from one of 55 amino acid sequences 42-876
 CC residues in length, corresponding to FLEXHT-1 to FLEXHT-55, a 90 %
 CC identical sequence, and a biologically active or immunogenic fragment of
 CC the sequence. The FLEXHT proteins can have antianaemic, anticonvulsant,
 CC antiarteriosclerotic, immunomodulatory, cytostatic, antiasthmatic,
 CC antiinflammatory, hepatotropic, antidiabetic, anti-gout, antihypertic,
 CC neuroprotective, antiallergic, osteopathic, antipsoriatic, antitumor
 CC and antirheumatic activities, and can be used in gene therapy. The
 CC polynucleotide sequences can be used to express the protein sequences.
 CC Pharmaceutical compositions comprising FLEXHT can be used to treat
 CC diseases or conditions associated with altered expression of functional
 CC FLEXHT. The proteins and polynucleotides can be used to diagnose and
 CC treat disorders including anaemia, epilepsy, arteriosclerosis,
 CC atherosclerosis, developmental disorders, cancers, and immunological
 CC disorders such as asthma, bronchitis, cirrhosis, Crohn's disease,
 CC diabetes mellitus, gout, Grave's disease, multiple sclerosis,
 CC osteoarthritis, pancreatitis, psoriasis, rheumatoid arthritis, and
 CC ulcerative colitis.
 XX
 XX Sequence 2273 BP; 494 A; 630 C; 547 G; 602 T; 0 other;
 SQ
 Query Match 100.0%; Score 2273; DB 22; Length 2273;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 421 TCACCTACCTTTGCTTCCGATGAGCTTCTGCTGATATTAATACCTGCCACCG 480
 QY
 DB 481 GCAACTACGATATCAAGTGGACATGTCATTTGTTCTGACTTTGAAGCTGATTTGTT 540
 QY
 DB 481 GCAACTACGATATCAAGTGGACATGTCATTTGTTCTGACTTTGAAGCTGATTTGTT 540
 QY
 DB 541 TGGCTGTTGACTTACTTTGACGAGGAGAAAGATCAGATTTCTTGTCTCTGAGCAACA 600
 DB 541 TGGCTGTTGACTTACTTTGACGAGGAGAAAGATCAGATTTCTTGTCTCTGAGCAACA 600
 QY
 DB 601 AATATGCACTAGTGTGTTCTTCCCTGCTGAGAAAGTGTGCTTCTCTGCTTCTGAG 660
 DB 601 AATATGCACTAGTGTGTTCTTCCCTGCTGAGAAAGTGTGCTTCTCTGCTTCTGAG 660
 QY
 DB 661 GGGCCCTTCTGTTAGAGGCCCACTTCTCATATGATCATCATGAGCTGTGACAGGAG 720
 DB 661 GGGCCCTTCTGTTAGAGGCCCACTTCTCATATGATCATCATGAGCTGTGACAGGAG 720
 QY
 DB 721 AGCTGATTGACATACCAAGAAAGATACCAACAGATCATTTCTGCTCTCAAGCCCTGA 780
 DB 721 AGCTGATTGACATACCAAGAAAGATACCAACAGATCATTTCTGCTCTCAAGCCCTGA 780
 QY
 DB 781 GTCTGAGGCTTTTCTACCTAGTGGGCTACACCTGCTAGCCGCCCATCACAGAACT 840
 DB 781 GTCTGAGGCTTTTCTACCTAGTGGGCTACACCTGCTAGCCGCCCATCACAGAACT 840
 QY
 DB 841 ATCTCTCACTAGAGATGACCAACCAACCCCTTGTGTTCCGCTGATGATCATGCTGA 900
 DB 841 ATCTCTCACTAGAGATGACCAACCAACCCCTTGTGTTCCGCTGATGATCATGCTGA 900
 QY
 DB 901 TCTGGGGCAAGTTTGTGCTGTCAAAATATGATCACTTGTGCTGCTCAAGAAAGAT 960
 DB 901 TCTGGGGCAAGTTTGTGCTGTCAAAATATGATCACTTGTGCTGCTCAAGAAAGAT 960
 QY
 DB 961 GCATTTTGAAGGAGGCTGCTCAATGCTTGAAGAAAGGCAAGCAAGTGGAGAT 1020
 DB 961 GCATTTTGAAGGAGGCTGCTCAATGCTTGAAGAAAGGCAAGCAAGTGGAGAT 1020
 QY
 DB 1021 CCTGTGCCAATGAAAGTGTGCTCTTGAAGAAACCCCGCTTCACTGAGCAACATTTG 1080
 DB 1021 CCTGTGCCAATGAAAGTGTGCTCTTGAAGAAACCCCGCTTCACTGAGCAACATTTG 1080
 QY
 DB 1081 CCTGATTTCAACATCAACCAACGCTGTGGTGGCCGCTACATCTTCAACAGACTGAAT 1140
 DB 1081 CCTGATTTCAACATCAACCAACGCTGTGGTGGCCGCTACATCTTCAACAGACTGAAT 1140
 QY
 DB 1141 TCTTGGAAATTAAGAACTCTCAAGGATCTCTGCTGCTATTTCTGAGCCCTCTGACAG 1200
 DB 1141 TCTTGGAAATTAAGAACTCTCAAGGATCTCTGCTGCTATTTCTGAGCCCTCTGACAG 1200
 QY
 DB 1201 GCTGTGACTCAGGATACCTGCTGTGCTTCAAGTGAATTCCTCATTTGTTATTTGAGAA 1260
 DB 1201 GCTGTGACTCAGGATACCTGCTGTGCTTCAAGTGAATTCCTCATTTGTTATTTGAGAA 1260
 QY
 DB 1261 GACAGGCTGCCAGGCTCATTTCAAGAGAGCCCACTGAGCAAGCTGGCCGCTATTACTG 1320
 DB 1261 GACAGGCTGCCAGGCTCATTTCAAGAGAGCCCACTGAGCAAGCTGGCCGCTATTACTG 1320
 QY
 DB 1321 TCTCTCAGGCTTCTTCAATTTGTTGTCAGACAGATTCATGAGCTCTTCAATGAGTTACT 1380
 DB 1321 TCTCTCAGGCTTCTTCAATTTGTTGTCAGACAGATTCATGAGCTCTTCAATGAGTTACT 1380
 QY
 DB 1381 CCAATGACTGCTTGTGCTCTTCACTGAGTGGAGAAATGAGTTAAGTTAATTCATCT 1440
 DB 1381 CCAATGACTGCTTGTGCTCTTCACTGAGTGGAGAAATGAGTTAAGTTAATTCATCT 1440
 QY
 DB 1441 ATTTCTTGGCCATCTTCTTCTGAGGCTTCAATTTATTTGCTTATTTCAAAAG 1500
 DB 1441 ATTTCTTGGCCATCTTCTTCTGAGGCTTCAATTTATTTGCTTATTTCAAAAG 1500
 QY
 DB 1501 CAATGTTGCCAAGAAAGAGTTAAAGAAATGAATTCATTTCTCTGTTGAGCT 1560
 DB 1501 CAATGTTGCCAAGAAAGAGTTAAAGAAATGAATTCATTTCTCTGTTGAGCT 1560

D	b		1501	CAATGAGTCCAAAGAAAAGAAGTTAAAGAAAGATGAATAATTCATTCTCCCTGGAGCCT	1560
Q	y		1561	GTCGGGGACTGTGTGCAGAAACTACTCGTCTCCCTTTTACACAGACTCCTTTGGCCCAGAG	1620
D	b		1561	GTCGGGGACTGTGTGCAGAAACTACTCGTCTCCCTTTTACACAGACTCCTTTGGCCCAGAG	1620
Q	y		1621	CAGAGAAATGAAAAAGCCAGGGAGGTGSAAGATCGATGCTTCAGCTGAGGCCCTGTGCC	1680
D	b		1621	CAGAGAAATGAAAAAGCCAGGGAGGTGSAAGATCGATGCTTCAGCTGAGGCCCTGTGCC	1680
Q	y		1681	AGCCAAGTCTTCATTTGGGGCCMAAGGGGAAACTTTTTTTTGAGAAAGCCCTCTTGCTTT	1740
D	b		1681	AGCCAAGTCTTCATTTGGGGCCMAAGGGGAAACTTTTTTTTGAGAAAGCCCTCTTGCTTT	1740
Q	y		1741	GTCACCCACGCTGGAATCGAGTGGCGGATCTCAGCTCACCGCAACTTCACCTTCCTGGG	1800
D	b		1741	GTCACCCACGCTGGAATCGAGTGGCGGATCTCAGCTCACCGCAACTTCACCTTCCTGGG	1800
Q	y		1801	TTCAGATTAATTTCCCTGCTCAGCCCTCCCAAGTAGCTGGAAATACGGCAGCCACCATG	1860
D	b		1801	TTCAGATTAATTTCCCTGCTCAGCCCTCCCAAGTAGCTGGAAATACGGCAGCCACCATG	1860
Q	y		1861	CCGAGCTAATTTTGTATTTTCAGTAGAAACGGGATTCACACAGTTGGCCAGGCTGTCTC	1920
D	b		1861	CCGAGCTAATTTTGTATTTTCAGTAGAAACGGGATTCACACAGTTGGCCAGGCTGTCTC	1920
Q	y		1921	TCGAATCTCTGACCGCAAGTAGATCCACCCTCCGCTCCCAAGTAGCTGGATTAACAG	1980
D	b		1921	TCGAATCTCTGACCGCAAGTAGATCCACCCTCCGCTCCCAAGTAGCTGGATTAACAG	1980
Q	y		1981	GCGTAGAGCAACGTGCCCCCGGCCCAAAGGGGAAACTCTTGTGGAGAGACAGAGGGGCTCA	2040
D	b		1981	GCGTAGAGCAACGTGCCCCCGGCCCAAAGGGGAAACTCTTGTGGAGAGACAGAGGGGCTCA	2040
Q	y		2041	CATCTCCCCCTGTGATTTCCCCCATGACATTTGCTTTATCTCTCCCATTTAGCCAGAAATC	2100
D	b		2041	CATCTCCCCCTGTGATTTCCCCCATGACATTTGCTTTATCTCTCCCATTTAGCCAGAAATC	2100
Q	y		2101	TATGTGTGTTTTCTTCTGSCCAATTACTATGATGTGTGATGTGCGCTACACACACCCCC	2160
D	b		2101	TATGTGTGTTTTCTTCTGSCCAATTACTATGATGTGTGATGTGCGCTACACACACCCCC	2160
Q	y		2161	CCCATGTGGGGGGGTGAGAGGGGTGCAAGGCCCTGCTGCTCACCTTTTCTAACCTTGGAA	2220
D	b		2161	CCCATGTGGGGGGGTGAGAGGGGTGCAAGGCCCTGCTGCTCACCTTTTCTAACCTTGGAA	2220
Q	y		2221	CTGATATTGATAAATCACTCTGTGTTGTAGCTTTCAAAAAAAAAAAAAAAAAA	2273
D	b		2221	CTGATATTGATAAATCACTCTGTGTTGTAGCTTTCAAAAAAAAAAAAAAAAAA	2273

RESULT 2	
AAF25169	
ID	AAF25169 standard; cDNA; 2264 BP.
XX	
AC	AAF25169;
XX	
DT	30-Apr-2001 (first entry)
XX	
DE	Nucleotide sequence of a human protein having a hydrophobic domain.
XX	
KW	Human; hydrophobic protein; secretory protein; membrane protein; sepsis;
KW	tumour inhibition; immune deficiency; autoimmune disorder; anaemia; burn;
KW	infectious disease; cancer; ulcer; periodontal disease; coagulation;
KW	Parkinson's disease; fertility; immune response; thrombosis; ss.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Location/Qualifiers
CDS	85..1548
FT	/*tag= a
FT	/product= "hydrophobic protein"
XX	

PN	WM0200104297-A2.
XX	18-JAN-2001.
FD	
XX	
PF	16-JUN-2000; 2000WO-JP03942.
XX	
PR	08-JUL-1999; 99JP-0194359.
XX	
PA	(SAGA) SAGAMI CHEM RES CENT.
PA	(PROT-) PROTEGENE INC.
PI	Kato S, Kimura T;
XX	
DR	WPI; 2001-103081/11.
DR	P-PSDB; AABJ31669.
XX	
PT	Isolated human proteins and polynucleotides are used in research and have activities including cell proliferation/differentiation activity,
PT	immune stimulating activity and receptor/ligand activity -
PS	Claim 4; Page 120-124; 151pp; English.
XX	
CC	The present sequence encodes a human protein with hydrophobic domains.
CC	AAR25159 represents a shorter version of the present sequence. The
CC	protein possesses a hydrophobic domain and so is a secretory protein
CC	or a membrane protein. The protein is used as an antigen to prepare
CC	antibodies. The polynucleotide sequence is useful as a source of probes
CC	for genetic diagnosis. It is also useful for producing the protein
CC	in large quantities and for gene therapy. The eukaryotic cells are used
CC	for detecting the receptors or ligands corresponding to the protein and
CC	for detecting small novel pharmaceuticals. The antibodies are also used
CC	for detection, quantification and purification of the proteins. Both the
CC	protein and polynucleotide may be used in research or as nutritional
CC	sources or supplements. The protein may have cytokine and cell
CC	proliferation/differentiation activity, immune stimulating or suppressing
CC	activity, hematopoiesis regulating activity, tissue growth activity,
CC	activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic
CC	and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC	activity and tumour inhibition activity. It may therefore may be used to
CC	treat immune deficiencies resulting from autoimmune disorders or
CC	infectious diseases, cancer, sepsis, anaemias, burns and ulcers.
CC	periodontal disease, Parkinson's disease, induce fertility, improve
CC	immune response and enhance coagulation or inhibit thrombolysis.
XX	
SQ	Sequence 2264 BP; 482 A; 629 C; 550 G; 603 T; 0 other;
Query Match	98.6%; Score 2241.2; DB 22; Length 2264;
Best Local Similarity	99.8%; Pred. No. 0; Mismatches 3; Indels 1; Gaps 1
Matches 2254; Conservative	0; Mismatches 3; Indels 1; Gaps 1
OY	1 GGGGGTGAGCGAATG-CGTTTGGCCCGCATTCGGGGCGCGGGAATGGGGGTCTCCTGT 59
Db	7 GGGGGTAGAGCAATGCGTTTGGCCCGCATTCGGGGCGCGGGAATGGGGGTCTCCTGT 66
OY	60 GGGGCTCCCGAGTTAAAGATGGCGTCTCAGCGGAGGGGACGAGGGGACTGTGTGGCG 119
Db	67 GGGGCTCCCGAGTTAAAGATGGCGTCTCAGCGGAGGGGACGAGGGGACTGTGTGGCG 126
OY	120 CTGGCGGGGGGTTCTGCAGTGGGGTTTCCAGAGCTGAGACCTTAACAATTGGCGAGTCC 179
Db	127 CTGGCGGGGGGTTCTGCAGTGGGGTTTCCAGAGCTGAGACCTTAACAATTGGCGAGTCC 186
OY	180 CTGGGCGCGTCAGAACAGGCGCTGCGGTGATCATCTCATCTTCTGGGTTAACCCCTTT 239
Db	187 CTGGGCGCGTCAGAACAGGCGCTGCGGTGATCATCTCATCTTCTGGGTTAACCCCTTT 246
OY	240 GCCTTGTATTATGGGCATTTACCTTTTCTACAAGAGACCTTACCTTACCTCTTTCAT 299
Db	247 GCCTTGTATTATGGGCATTTACCTTTTCTACAAGAGACCTTACCTTACCTCTTTCAT 306
OY	300 ACCCTTACAGGCGCTCAATGCTATTTTAACTTTGAAAACGAGCTTACACATCCCTG 359
Db	307 ACCCTTACAGGCGCTCAATGCTATTTTAACTTTGAAAACGAGCTTACACATCCCTG 366

Db 1320 TACTATTGGTGCACACACATCCACTGGCTTATAGGTTACTCCATGACTGCTTC 1379
QY 1395 TGGCTCTTCACTGGGACCAATGGCTTAAGTGTATAAATCCATTAATTCCTTGGCCAC 1454
Db 1380 TGGCTCTTCACTGGGACCAATGGCTTAAGTGTATAAATCCATTAATTCCTTGGCCAC 1439
QY 1455 ATCTTCTTCTGAGCTTACTTATTCATTTGCTTATTTTCAACAAGCATGTCCTCAAG 1514
Db 1440 ATCTTCTTCTGAGCTTACTTATTCATTTGCTTATTTTCAACAAGCATGTCCTCAAG 1499
QY 1515 AAGAGAGATTAAAGAGATGATATATCCATTTCCCTGGTGGCTGTGGGAGATGCTGTG 1574
Db 1500 AAGAGAGATTAAAGAGATGATATATCCATTTCCCTGGTGGCTGTGGGAGATGCTGTG 1559
QY 1575 CAGAACTACTGCTCTCTCTTTCACAGCACTCTCTTGGCCAGAGAGAAATGAGAAA 1634
Db 1560 CAGAACTACTGCTCTCTCTTTCACAGCACTCTCTTGGCCAGAGAGAAATGAGAAA 1619
QY 1635 GCCAGGAGGTGGAAGATGATGCTTCCAGCTGTGCTGTGCTCCAGCAATCTTCTCAT 1694
Db 1620 GCCAGGAGGTGGAAGATGATGCTTCCAGCTGTGCTGTGCTCCAGCAATCTTCTCAT 1679
QY 1695 TTGGGGCCAAAGGGGAACTTTTCTTGGAGAAAGGCTTGTGCTTGTCCACCCAGCTGG 1754
Db 1680 TTGGGGCCAAAGGGGAACTTTTCTTGGAGAAAGGCTTGTGCTTGTCCACCCAGCTGG 1739
QY 1755 AATGAGTGGCGGGGATCTCAGCTCACCGCAACCTCCACTCTGGTTCAAGTATTTTC 1814
Db 1740 AATGAGTGGCGGGGATCTCAGCTCACCGCAACCTCCACTCTGGTTCAAGTATTTTC 1799
QY 1815 CTGCTTCAAGCTCCCAAGTATGCTGGAAATACAGGACGCAATGCGCCAGCTAATTTT 1874
Db 1800 CTGCTTCAAGCTCCCAAGTATGCTGGAAATACAGGACGCAATGCGCCAGCTAATTTT 1859
QY 1875 GATATTTTCAAGTATGAGGATTTTACACGATTTGGCCAGGCTGTCTGAACTCTGACC 1934
Db 1860 GATATTTTCAAGTATGAGGATTTTACACGATTTGGCCAGGCTGTCTGAACTCTGACC 1919
QY 1935 GCAAGTATCCACCCGCTCTCCGCTCCCAAGTATGCTGGAAATACAGGCTGTGACCCT 1994
Db 1920 GCAAGTATCCACCCGCTCTCCGCTCCCAAGTATGCTGGAAATACAGGCTGTGACCCT 1979
QY 1995 GCCCGGCCCAAGGGGAACTCTTGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2054
Db 1980 GCCCGGCCCAAGGGGAACTCTTGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2039
QY 2055 TTCCCCCATGACATGCTTATCTCTCCCATCTAGCAGGAATCTATTGTGTTTCT 2114
Db 2040 TTCCCCCATGACATGCTTATCTCTCCCATCTAGCAGGAATCTATTGTGTTTCT 2099
QY 2115 TCTGCCAATTTACTATGATTTGTATGTGCTTACCAACACCCCCCATGAGGGGGGTG 2174
Db 2100 TCTGCCAATTTACTATGATTTGTATGTGCTTACCAACACCCCCCATGAGGGGGGTG 2159
QY 2175 GAGAGGGGTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2234
Db 2160 GAGAGGGGTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2219
QY 2235 ATCACTTCTGTGTTGT 2249
Db 2220 GTCACTTCTGTGTTGT 2234

RESULT 4
ABK84420
ID ABK84420 standard; cDNA; 1842 BP.
XX
AC ABK84420;
XX
XX 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #991.
XX

KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
PN MO20028999-A2.
PD 11-APR-2002.
PF 03-OCT-2001; 2001WO-US30821.
PR 03-OCT-2000; 2000US-237189P.
XX (GENE-) GENE LOGIC INC.
PA Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
PI WPI; 2002-435328/46.
DR
XX
XX
PT Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity -
XX
PS Claim 1; SEQ ID No 991; 114pp; English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and MS is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1842 BP; 413 A; 525 C; 413 G; 491 T; 0 other;
SQ

Query Match 81.0%; Score 1842; DB 24; Length 1842;
Best Local Similarity 100.0%; Pred. No. 0;

Db 1561 GGAAATACAGCAGCCACCATGCCCAGTAAATTTTGTATTTTTCAGTAGMAAGGGGATTT 1620
OY 1899 CACCACTGTGGCAGGCTGATCTCGAATCTCTGACCGCAAGTATCCACCGCTCCGCC 1958
Db 1621 CACCACTGTGGCAGGCTGATCTCGAATCTCTGACCGCAAGTATCCACCGCTCCGCC 1680
OY 1959 TCCCAAGTGTGGGATTTACAGGCGTGAGCCACCGTCCCGGCCCAAGGGGAACTCTT 2018
Db 1681 TCCCAAGTGTGGGATTTACAGGCGTGAGCCACCGTCCCGGCCCAAGGGGAACTCTT 1740
OY 2019 GTGGAGAGGAGCAGAGGGGCTCACATCTCCCTCTGATTTCCCGCATGCACTTGCTTATC 2078
Db 1741 GTGGAGAGGAGCAGAGGGGCTCACATCTCCCTCTGATTTCCCGCATGCACTTGCTTATC 1800
OY 2079 TCTCCCATCTAGCCAGCAATCTATTTGTTTCTTCTGCC 2120
Db 1801 TCTCCCATCTAGCCAGCAATCTATTTGTTTCTTCTGCC 1842
RESULT 6
ABL64098
ID ABL64098 standard; DNA; 1842 BP.
XX
AC ABL64098;
XX
DT 15-MAY-2002 (first entry)
XX
DE Breast cancer related gene sequence SEQ ID NO:2435.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
KW gene; de.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209511P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.

PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 2435; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (1) of a signature gene set, where (1)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (1) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.
XX
SQ Sequence 1842 BP; 413 A; 525 C; 413 G; 491 T; 0 other;
Query Match 81.0%; Score 1842; DB 24; Length 1842;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 279 TACCTCATCCACCTTCCATACCTTTACAGGCGCTCAATGCTTATTTTAACTTGA 338
Db 1 TACCTCATCCACCTTCCATACCTTTACAGGCGCTCAATGCTTATTTTAACTTGA 60
OY 339 AACCACTTACCACTCCCTGCTGTGATTTGCTTCACTTCCATCTTGACTAATG 398
Db 61 AACCACTTACCACTCCCTGCTGTGATTTGCTTCACTTCCATCTTGACTAATG 120
OY 399 GGGCGCACCATCTGCGGTCTCTACTACTTTTCTTCAGATGGCTTACTTGTGCT 458
Db 121 GGGCGCACCATCTGCGGTCTCTACTACTTTTCTTCAGATGGCTTACTTGTGCT 180
OY 459 GGATACCTTACAGTCCAGCGGCACTACATCAAGTGGAGCAATGCATATGTGT 518
Db 181 GGATACCTTACAGTCCAGCGGCACTACATCAAGTGGAGCAATGCATATGTGT 240
OY 519 CTGACTTTGAAGCTGATTTGCTTGTGCTTGACTTCTTGAAGGAGGAATGAGAT 578
Db 241 CTGACTTTGAAGCTGATTTGCTTGTGCTTGACTTCTTGAAGGAGGAATGAGAT 300
OY 579 TCTTGTCTCTGAGCAACAGAAATATGCACTAGCTGTGTTCTTCTGCTGGAATT 638
Db 301 TCTTGTCTCTGAGCAACAGAAATATGCACTAGCTGTGTTCTTCTGCTGGAATT 360

DB 1141 CTGGCTGCTTCAGATGGAATTCCTCATGTTATGGAAGACAGGCTGCAGAGCTC 1200
QY 1278 ATTCAGAGAGGCCCCCCTGAGCAGCTGGCCGCATACCTCTCTCCAGCCCTTTC 1337
DB 1201 ATTCAGAGAGGCCCCCCTGAGCAGCTGGCCGCATACCTCTCTCCAGCCCTTTC 1260
QY 1338 TATTGTGTCACAGACATCCATCTGCTCTTCATGAGGTATCTCATGACTGCTTCTG 1397
DB 1261 TATTGTGTCACAGACATCCATCTGCTCTTCATGAGGTATCTCATGACTGCTTCTG 1320
QY 1398 CTCTTCACCTGGAGCAATAGCTTAAAGTGTATTAATCCATCTATTTCTTGCCACATC 1457
DB 1321 CTCTTCACCTGGAGCAATAGCTTAAAGTGTATTAATCCATCTATTTCTTGCCACATC 1380
QY 1458 TCTTCCTGAGCCTCTATTCATATTTGCTTATTCACAAAGCAATGCTGCCAAGAA 1517
DB 1381 TCTTCCTGAGCCTCTATTCATATTTGCTTATTCACAAAGCAATGCTGCCAAGAA 1440
QY 1518 GAGAGTTAAAGAGATGGA 1538
DB 1441 GAGAGTTAAAGAGATGGA 1461

RESULT 9
ABV30301
ID ABV30301 standard; cDNA; 1296 BP.
XX

ABV30301;

16-SEP-2002 (first entry)

Human prostate expression marker cDNA 30292.

Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
pharmacogenomic marker; gene; ss.

OS Homo sapiens.

PN W0200160860-A2.

23-AUG-2001.

20-FEB-2001; 2001WO-US05171.

17-FEB-2000; 2000US-183319P.

16-MAR-2000; 2000US-189862P.

25-MAY-2000; 2000US-207454P.

09-JUN-2000; 2000US-211314P.

18-JUL-2000; 2000US-219007P.

13-DEC-2000; 2000US-255281P.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Endege WO, Monahan JB;

WPI; 2001-662795/76.

Novel isolated nucleic acid molecule associated with cancerous state of

prostate cells and correlating with presence of prostate cancer, useful

for detecting presence of prostate cancer, stage of prostate cancer -

Claim 1; Page 6582; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (1) comprising

a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

specification or its complement. (1) is useful for:

(a) assessing whether a patient is afflicted with prostate cancer;

(b) monitoring the progression of prostate cancer in a patient;

(c) assessing the efficacy of a test compound to inhibit prostate

cancer in a patient;

(d) assessing the efficacy of a therapy for inhibiting prostate cancer

in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX

Sequence 1296 BP; 324 A; 342 C; 320 G; 296 T; 14 other;

Query Match 34.0%; Score 773.6; DB 23; Length 1296;

Best Local Similarity 95.9%; Pred. No. 5.7e-206;

Matches 794; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 957 GTATGATTTTACAGGCGCTGGCTTCAATGCTTTGAAGAAAGGCAAGGCAAGTGG 1016
DB 91 GTATGATTTTACAGGCGCTGGCTTCAATGCTTTGAAGAAAGGCAAGGCAAGTGG 150
QY 1017 GATGCTGTGCCAATGAGGTGTGCTCTTTGAACAAACCCCGCTTCACTGGCACC 1076
DB 151 GATGCTGTGCCAATGAGGTGTGCTCTTTGAACAAACCCCGCTTCACTGGCACC 210
QY 1077 ATTGCTCATTCACATCAACACCAAGCTGGGTGGCCCGCTTCACTTCAAGCACTC 1136
DB 211 ATGGCTCATTCACATCAACACCAAGCTGGGTGGCCCGCTTCACTTCAAGCACTC 270
QY 1137 AAGTCCCTTGAAGAAATGAAGAACTCTGAGGCTCTGCTGCTATTCCTGGCCCTTGG 1196
DB 271 AAGTCCCTTGAAGAAATGAAGAACTCTGAGGCTCTGCTGCTATTCCTGGCCCTTGG 330
QY 1197 CAGGCTGTGCACTCAGATACCTGTGCTTCCAGATGAATTCCTCATTTGATTTGG 1256
DB 331 CAGGCTGTGCACTCAGATACCTGTGCTTCCAGATGAATTCCTCATTTGATTTGG 390
QY 1257 GAAAGCAGGCTGCCAGGCTCTTCAAGAGAGCCCACTGAGCAAGCTGCGCCATT 1316
DB 391 GAAAGCAGGCTGCCAGGCTCTTCAAGAGAGCCCACTGAGCAAGCTGCGCCATT 450
QY 1317 ACTGCTCTCAGGCGCTTCAATGCTTGTGCAAGACACCATCAGCTGCTTCAATGGT 1376
DB 451 ACTGCTCTCAGGCGCTTCAATGCTTGTGCAAGACACCATCAGCTGCTTCAATGGT 510
QY 1377 TACTCCATGACTGCTTGTGCTCTTCACTGAGCAATGCTTAAAGTGTATAATTC 1436
DB 511 TACTCCATGACTGCTTGTGCTCTTCACTGAGCAATGCTTAAAGTGTATAATTC 570
QY 1437 ATCTATTTCTTGGGCAATCTTCTTCTGAGCTTACTATTAATGCTTAAATTC 1496
DB 571 ATCTATTTCTTGGGCAATCTTCTTCTGAGCTTACTATTAATGCTTAAATTC 630
QY 1497 AAGCAATGTCGCAAGAAAGAAAGAAAGTAAAGAAATGGAATATTCATTTCCCTGGTG 1556
DB 631 AAGCAATGTCGCAAGAAAGAAAGAAAGTAAAGAAATGGAATATTCATTTCCCTGGTG 690
QY 1557 GCTGTGGGAGAGTGTGTCAGAACTACTGCTCTCTTTTACAGACTCTTTGGCCC 1616
DB 691 GCTGTGGGAGAGTGTGTCAGAACTACTGCTCTCTTTTACAGACTCTTTGGCCC 750
QY 1617 AGAGCAGAGATGGAAGAAAGCAGGAGGTGAAGATGATGCTTCAAGCTGTGCTTGC 1676
DB 751 AGAGCAGAGATGGAAGAAAGCAGGAGGTGAAGATGATGCTTCAAGCTGTGCTTGC 810
QY 1677 TGCCAGCAAGTCTCATTTTGGGGCAAGAGGGAATCTTTTGGAGAAAGGCTCTTG 1736
DB 811 TGCCAGCAAGTCTCATTTTGGGGCAAGAGGGAATCTTTTGGAGAAAGGCTCTTG 870
QY 1737 CTCTGTCAACCAAGCTGGAATGAGTGGCGGATCTCAGTCAACCGCA 1784
DB 871 TTGTGTACCAACGCTGGAATGCACTGCGGATCTCAGTCAACCGCA 918

RESULT 10
AAS30639/C
ID AAS30639 standard; DNA; 7461 BP.

AC AAS30639;
XX
DT 21-NOV-2001 (first entry)
XX
DE DNA encoding novel lung cancer antigen, Seq ID No 91.
XX
KW Human; lung cancer; immunosuppressive; antiarthritic; antirheumatic;
KM antiproliferative; cytostatic; cardiac; vasoconstrictive; cerebroprotective;
KM neurotropic; neuroprotective; antibacterial; vitinoid; fungicide;
KM ophthalmological; autoimmune disease; rheumatoid arthritis; neoplasm;
KM hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KM nervous system disorder; Alzheimer's disease; infection; skin aging;
XX ocular disorder; wound healing; organ transplantation; ds.
OS Homo sapiens.
FN WO200155300-A2.
PN 02-AUG-2001.
PD 17-JAN-2001; 2001WO-USO1238.
PF 31-JAN-2000; 2000US-0179065.
PR (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-465565/50.
DR Isolated nucleic acid molecule encoding a lung cancer antigen is used
XX in preventing, treating or ameliorating a medical condition -
PS Disclosure; SEQ ID No 91; 475bp; English.

CC The invention relates to novel isolated lung cancer antigen
CC polynucleotides (I) and polypeptides (II). (I) and (II) are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. (I) and (II) are
CC also used in diagnosing a pathological condition or susceptibility to a
CC pathological condition, in particular, lung cancer. The antibodies to (I)
CC (II) can also be used in alleviating symptoms associated with the
CC disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme
CC linked immunoabsorbent assays (ELISA). Disorders which are diagnosed or
CC treated include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischemia, angiodementia, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
CC ocular disorders e.g. corneal infection. The polypeptides can also be
CC used to aid wound healing and epithelial cell proliferation, to prevent
CC skin aging due to sunburn, to maintain organs before transplantation, for
CC supporting cell culture of primary tissues, to regenerate tissues and in
CC chemotherapeutics. AAS30580-AAS30685 represent novel human lung cancer antigen
CC coding sequences, PCR primers and related sequences of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WHO at: ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 7461 BP; 1883 A; 1666 C; 1923 G; 1989 T; 0 other;

Query Match 31.2%; Score 708.6; DB 22; Length 7461;
Best Local Similarity 99.4%; Pred. No. 2.3e-187;
Matches 711; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1546 TTTCCTTGAGTGGCGGACTGTGCAGAAATCACTCCTTCACAGCAC 1605
DB 721 TCCTGCAAGTAGCTGCGGGAATGCTGCAAGAACTACTCCTTTACAGAC 662

OY 1606 TCCTTTGCCAGACAGAAATGAAAAGCAGGAGAGTGAAGATGATGCTTCAGC 1655
DB 661 TCCTTTGCCAGACAGAAATGAAAAGCAGGAGAGTGAAGATGATGCTTCAGC 602

Oy		1666	TGAGCCTCTGTCCACGCCAAGCTTCATTGGGGCCAAAGGGGAACCTTTTITGGAG	1725
Db		601	TGAGCCTCTGTCCACGCCAAGCTTCATTGGGGCCAAAGGGGAACCTTTTITGGAG	542
Oy		1726	AAGCGCTTGCTTTGTCAACCACGCCTGAGATCATGAGCGGATCTCACCTACCGCAA	1785
Db		541	AAGCGCTTGCTTTGTCAACCACGCCTGAGATCATGAGCGGATCTCACCTACCGCAA	482
Oy		1786	CCTCCACTCTCTGGGTCAAGTAATTTCTGTGCTCAGCCTCCCAATGACTGGGAATAC	1845
Db		481	CCTCCACTCTCTGGGTCAAGTAATTTCTGTGCTCAGCCTCCCAATGACTGGGAATAC	422
Oy		1846	AGGCACGCCACCATGCCCCAGCTAATTTTTGTATTTTCAGTAGAAAACGGATTTTCACCA	1905
Db		421	AGGCACGCCACCATGCCCCAGCTAATTTTTGTATTTTCAGTAGAAAACGGATTTTCACCA	362
Oy		1906	TTGGCCAGGCTGTGTTCTTGAACTCTTGACCGCAAGTAGATCCACCCGCTCCGCCAAA	1965
Db		361	TTGGCCAGGCTGTGTTCTTGAACTCTTGACCGCAAGTAGATCCACCCGCTCCGCCAAA	302
Oy		1966	GTCCTGGGATTAACAGCGTGTAGCACAACGTCGCGGCCCAAGGGGAAACCTGTGGGAG	2025
Db		301	GTCCTGGGATTAACAGCGTGTAGCACAACGTCGCGGCCCAAGGGGAAACCTGTGGGAG	242
Oy		2026	GAGCAGAGGGGCTCACATCTCCCCTGTGATTTCCCCCATGACATAATGCTTATCTCCCC	2085
Db		241	GAGCAGAGGGGCTCACATCTCCCCTGTGATTTCCCCCATGACATAATGCTTATCTCCCC	182
Oy		2086	ATCTAGCCAGGAATCTATTTGTGTTTTCTTCTGCAATTACTATGATTTGATGTGCC	2145
Db		181	ATCTAGCCAGGAATCTATTTGTGTTTTCTTCTGCAATTACTATGATTTGATGTGCC	122
Oy		2146	GCTACCAACACCCCCCATGGGGGGTGGAGAAGGGGTGCAAGGCCCTGCGCTGCCACT	2205
Db		121	GCTACCAACACCCCCCATGGGGGGTGGAGAAGGGGTGCAAGGCCCTGCGCTGCCACT	62
Oy		2206	TTTTCTACTTGGNACTGTATTAGATPMAATCACTTCTGTTTGTTCAGTTTTTCA	2260
Db		61	TTTTCTACTTGGNACTGTATTAGATPMAATCACTTCTGTTTGTTCAGTTTTTCA	7
 RESULT 11 AAS28701/c ID AAS28701 standard; DNA; 7461 BP.				
XX		AAS28701:		
AC				
XX				
DT		07-NOV-2001 (first entry)		
XX				
DE		Genomic sequence #541 encoding for novel human respiratory antigen.		
XX				
KM		Human, respiratory antigen; respiratory disorder; throat disorder;		
KW		lung disorder; nose disorder; lung cancer; gene therapy; cytosolic;		
KM		anti allelogic; anti ashmatic; anti inflammatory; olfactory;		
KW		respiratory active; ds.		
XX				
OS		Homo sapiens.		
XX				
PN		WO200155448-A1.		
XX				
PD		02-AUG-2001.		
XX				
PF		17-JAN-2001; 2001WO-US01333.		
XX				
PR		31-JAN-2000; 2000US-0179065.		
PR		04-FEB-2000; 2000US-0180628.		
PR		24-FEB-2000; 2000US-0184664.		
PR		02-MAR-2000; 2000US-0186350.		
PR		16-MAR-2000; 2000US-0189874.		
PR		17-MAR-2000; 2000US-0190076.		
PR		18-APR-2000; 2000US-0198123.		
PR		19-MAY-2000; 2000US-0205515.		

PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 20-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCT INC.
Rosen CA, Barash SC, Ruben SM;
WPI, 2001-476224/51.
The present invention relates to the isolation of novel human
respiratory antigens (AAU17685-AAU17975), and cDNA and genomic
sequences encoding for these polypeptides. The sequences of the
invention are useful for preventing, treating and/or prognosing
disorders related to the respiratory system including respiratory
cancers and also for testing and detection e.g. diagnosis -
Isolated polypeptide for treating, preventing and/or prognosing
disorders related to the respiratory system including respiratory
cancers and also for testing and detection e.g. diagnosis -
Disclousure; SED ID No 1135; 546pp; English.

CC disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis),
CC lung disorders e.g. pneumonia, allergic disorders e.g. asthma,
CC pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of
CC the respiratory tissues e.g. lung cancer. The polynucleotide sequences
CC of the invention are useful in gene therapy and antisense therapy.
CC AAS28161-AAS28764 represent genomic sequences encoding for novel
CC human respiratory antigens.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7461 BP; 1883 A; 1666 C; 1923 G; 1989 T; 0 other;
Query Match 31.2%; Score 708.6; DB 22; Length 7461;
Best Local Similarity 99.4%; Pired. No. 2,3e-187;
Matches 711; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1546 TTTCCTGGTGGCTGCTGGGAGCTGGTGCAAACTACTGCTCCCTTTTACAGAC 1605
DB 721 TCTTGCAGGTGGCTGCTGGGAGCTGGTGCAAACTACTGCTCCCTTTTACAGAC 662
QY 1606 TCCTTTGCCCCAGAGCAGAGAAATGAAAAGCCAGGAGGTGAAATGATGCTTCAGC 1665
DB 661 TCCTTTGCCCCAGAGCAGAGAAATGAAAAGCCAGGAGGTGAAATGATGCTTCAGC 602
QY 1666 TGTGCTCTGTCTGCTCCAGCCAGCTTCATTTGGGGCCAAAGGGGAACTTTTGGAG 1725
DB 601 TGTGCTCTGTCTGCTCCAGCCAGCTTCATTTGGGGCCAAAGGGGAACTTTTGGAG 542
QY 1726 AAGGCGTCTGCTTTTGTGACCCAGCTGGAATGAGTGGGGGATCTCAGCTCAGCGCA 1785
DB 541 AAGGCGTCTGCTTTTGTGACCCAGCTGGAATGAGTGGGGGATCTCAGCTCAGCGCA 482
QY 1786 CCTCAGCTCTGGGTTTCAAGTATTTCTGCTCAGCTCCCAAGTACGGGAATAC 1845
DB 481 CCTCAGCTCTGGGTTTCAAGTATTTCTGCTCAGCTCCCAAGTACGGGAATAC 422
QY 1846 AGGCAAGCCCACTGCCCACTAATTTTGTATTTTCAAGTAAAGGGAATTTACACAG 1905
DB 421 AGGCAAGCCCACTGCCCACTAATTTTGTATTTTCAAGTAAAGGGAATTTACACAG 362
QY 1906 TTGGCAGGCTGTCTGTAATCTCTGACCGCAAGTATCACCAGGCTCCGCTCCCAA 1965
DB 361 TTGGCAGGCTGTCTGTAATCTCTGACCGCAAGTATCACCAGGCTCCGCTCCCAA 302
QY 1966 GTGCTGGGATTACAGGGGTGAGCCACCGTCCCGGCCAAAGGGGAACTTTGGGAG 2025
DB 301 GTGCTGGGATTACAGGGGTGAGCCACCGTCCCGGCCAAAGGGGAACTTTGGGAG 242
QY 2026 GAGCAGAGGGGCTCAGATCTCCCTCGATTTCCCAATGACATGCTTATCTTCC 2085
DB 241 GAGCAGAGGGGCTCAGATCTCCCTCGATTTCCCAATGACATGCTTATCTTCC 182
QY 2086 ATCTAGCAGGAATCTATTTGTTTCTTCTGCAATTTACTATGATGATGAGCC 2145
DB 181 ATCTAGCAGGAATCTATTTGTTTCTTCTGCAATTTACTATGATGATGAGCC 122
QY 2146 GCTACCAACACCCCCCATGGGGGGTGAAGAGGGGTGCAAGGCCCTGCTCTCACT 2205
DB 121 GCTACCAACACCCCCCATGGGGGGTGAAGAGGGGTGCAAGGCCCTGCTCTCACT 62
QY 2206 TTTTCTAAGCTTGAATCTATTTAGTAAATCACTTGTGTTGTGCTTTTCA 2260
DB 61 TTTTCTAAGCTTGAATCTATTTAGTAAATCACTTGTGTTGTGCTTTTCA 7
RESULT 12
ID ACA03402/c standard; DNA; 7461 BP.
XX ACA03402;
XX
XX 22-MAY-2003 (first entry)

XX DE DNA encoding human lung cancer antigen HIPA05.
XX
XX Human; ds; gene; lung cancer antigen; anaemia; tissue regeneration;
XX constant region; immunotherapy; gene therapy; lung disorder; leukopenia;
XX small cell lung cancer; squamous cell carcinoma; adenocarcinoma; asthma;
XX respiratory disorder; nonallergic rhinitis; rheumatoid arthritis; ulcer;
XX adult respiratory distress syndrome; hyperproliferative disorder; graft;
XX lung neoplasm; prostate neoplasm; immune system disorder; anaphylaxis;
XX multiple sclerosis; blood-related disorder; allergic reaction; impotence;
XX inflammatory disorder; appendicitis; dermatitis; graft vs host disease;
XX immune complex disease; serum sickness; polyarteritis nodosa; arhythmia;
XX urinary system disorder; glomerulonephritis; kidney failure; infertility;
XX cardiovascular disorder; myocardial infarction; osteoarthritis; aging;
XX musculoskeletal system disorder; Albers-Schönberg disease; cancer; burn;
XX neurological disorder; Alzheimer's disease; Parkinson's disease; trauma;
XX endocrine disorder; Addison's disease; diabetes mellitus; gastric reflux;
XX gastrointestinal disorder; reproductive system disorder; cerebral palsy;
XX developmental disorder; Panconi's syndrome; cellular level disease;
XX amyotrophic lateral sclerosis; infectious disease; viral infection;
XX bacterial infection; wound healing.
XX
XX Homo sapiens.
XX
XX US2002173454-A1.
XX
XX 21-NOV-2002.
XX
XX 17-JAN-2001; 2001US-0764904.
XX
XX 31-JUN-2000; 2000US-179065P.
XX 04-FEB-2000; 2000US-180628P.
XX 28-JUN-2000; 2000US-214886P.
XX 07-JUL-2000; 2000US-216647P.
XX 07-JUL-2000; 2000US-216880P.
XX 11-JUL-2000; 2000US-217487P.
XX 11-JUL-2000; 2000US-217496P.
XX 14-JUL-2000; 2000US-218290P.
XX 26-JUL-2000; 2000US-220963P.
XX 26-JUL-2000; 2000US-220964P.
XX 14-AUG-2000; 2000US-224518P.
XX 14-AUG-2000; 2000US-224519P.
XX 14-AUG-2000; 2000US-225267P.
XX 14-AUG-2000; 2000US-225268P.
XX 14-AUG-2000; 2000US-225270P.
XX 14-AUG-2000; 2000US-225276P.
XX 14-AUG-2000; 2000US-225447P.
XX 14-AUG-2000; 2000US-225757P.
XX 14-AUG-2000; 2000US-225758P.
XX 22-AUG-2000; 2000US-226868P.
XX 30-AUG-2000; 2000US-228924P.
XX 01-SEP-2000; 2000US-229287P.
XX 01-SEP-2000; 2000US-229343P.
XX 01-SEP-2000; 2000US-229344P.
XX 01-SEP-2000; 2000US-229345P.
XX 05-SEP-2000; 2000US-229509P.
XX 05-SEP-2000; 2000US-229513P.
XX 08-SEP-2000; 2000US-231413P.
XX 21-SEP-2000; 2000US-234223P.
XX 21-SEP-2000; 2000US-234274P.
XX 25-SEP-2000; 2000US-234997P.
XX 27-SEP-2000; 2000US-235834P.
XX 29-SEP-2000; 2000US-236327P.
XX 29-SEP-2000; 2000US-236367P.
XX 29-SEP-2000; 2000US-236368P.
XX 29-SEP-2000; 2000US-236369P.
XX 29-SEP-2000; 2000US-236370P.
XX 02-OCT-2000; 2000US-236802P.
XX 02-OCT-2000; 2000US-237037P.
XX 02-OCT-2000; 2000US-237038P.
XX 02-OCT-2000; 2000US-237039P.
XX 02-OCT-2000; 2000US-237040P.
XX 13-OCT-2000; 2000US-239935P.
XX 20-OCT-2000; 2000US-240960P.

PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-020515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234957.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236337.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244624.
PR 08-NOV-2000; 2000US-0244625.
PR 08-NOV-2000; 2000US-0244626.
PR 08-NOV-2000; 2000US-0244627.
PR 08-NOV-2000; 2000US-0244628.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX
XX Rosen CA, Barash SC, Ruben SM,
PI
XX
XX MPI; 2001-541565/60.
XX
XX Nucleic acids encoding 324 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
XX cancers and metastases -
XX
XX Disclosure; SEQ ID NO 7996; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABA14678-ABA18001) useful for preventing, treating or ameliorating

PR	08-SEP-2000	2000US-02311414
PR	08-SEP-2000	2000US-0232080
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0232397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232400
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234224
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	25-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0235327
PR	29-SEP-2000	2000US-0235367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236370
PR	29-SEP-2000	2000US-0236379
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239935
PR	13-OCT-2000	2000US-0239937
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241826
PR	20-OCT-2000	2000US-0242221
PR	01-NOV-2000	2000US-0246174
PR	01-NOV-2000	2000US-0246175
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246539
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249208
PR	17-NOV-2000	2000US-0249209
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249212
PR	17-NOV-2000	2000US-0249213
PR	17-NOV-2000	2000US-0249214
PR	17-NOV-2000	2000US-0249215
PR	17-NOV-2000	2000US-0249216
PR	17-NOV-2000	2000US-0249217
PR	17-NOV-2000	2000US-0249218
PR	17-NOV-2000	2000US-0249244
PR	17-NOV-2000	2000US-0249245
PR	17-NOV-2000	2000US-0249246
PR	17-NOV-2000	2000US-0249257
PR	17-NOV-2000	2000US-0249265
PR	17-NOV-2000	2000US-0249295

PR	17-NOV-2000;	2000US-0249329.
PR	17-NOV-2000;	2000US-0249330.
PR	01-DEC-2000;	2000US-0250391.
PR	01-DEC-2000;	2000US-0250391.
PR	01-DEC-2000;	2000US-0251160.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
XX	Rosen CA, Barash SC, Ruben SM;	
PI		
XX	WPI, 2001-541565/60.	
XX		
DR		
PT	Nucleic acids encoding 3224 human nervous system antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating nervous system	
PT	cancers and metastases -	
PS	Disclosure; SEQ ID NO 11808; 1701bp + Sequence Listing; English.	
XX		
XX	The invention relates to novel genes (ABAI1004-ABA21534) and proteins	
CC	(ABAI14678-ABAI18001) useful for preventing, treating or ameliorating	
CC	medical conditions e.g. by protein or gene therapy. The genes are	
CC	isolated from a range of human tissues disclosed in the specification.	
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful	
CC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast	
CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone	
CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;	
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune	
CC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's	
CC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative	
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemia;	
CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and	
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal	
CC	and parasitic infections.	
CC	Note: The sequence data for this patent did not form part of the	
CC	printed specification, but was obtained in electronic format directly	
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
XX		
SQ	Sequence 32174 BP; 8747 A; 7064 C; 6837 G; 8001 T; 1525 other;	
Query Match	31.2%; Score 708.6; DB 22; Length 32174;	
Best Local Similarity	99.4%; Pred. No 5,1e-187;	
Matches	711; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
DG		
OY	1546 TTTCCTTGTTGGCCTGTGCGGGA CTGGTGCAGAAACTACTCGTCTCCCTTTACAGCAC	1605
DB	26125 TCTTGACAGGTGGCTGTGCGGGA CTGGTGCAGAAACTACTCGTCTCCCTTTACAGCAC	26184
OY	1606 TCCTTTGGCCCCAGAGCAGAGAATGAAAAAACCACGAGGAGGTGGAAGATCGATTGCTCCAAC	1665
DB	26185 TCCTTTGGCCCCAGAGCAGAGAATGAAAAAACCACGAGGAGGTGGAAGATCGATTGCTCCAAC	26244
OY	1666 TGTCGCTCTGTGTCAGCAGCAAGTCTTATTTTGAGGCCAAAAGSGGAATACTTTTTTTGAG	1725
DB	26245 TGTCGCTCTGTGTCAGCAGCAAGTCTTATTTTGAGGCCAAAAGSGGAATACTTTTTTTGAG	26304
OY	1726 AAGGCGTCTTGCTTTGTTCACCCACGCTGAATGACAGTGGCGGATCTCAGCTCACCGCAA	1785
DB	26305 AAGGCGTCTTGCTTTGTTCACCCACGCTGAATGACAGTGGCGGATCTCAGCTCACCGCAA	26364
OY	1786 CCTCACA CTCTGGAGTTCAAGTATTTTCTGTGCTCAGCCTCCCACAGTACGGGAATAC	1845
DB	26365 CCTCACA CTCTGGAGTTCAAGTATTTTCTGTGCTCAGCCTCCCACAGTACGGGAATAC	26424

PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI: 2001-541565/60.
 XX
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX
 PS Disclosure; SEQ ID NO 12680; 1701pp + Sequence Listing; English.

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 32174 BP; 8747 A; 7064 C; 6837 G; 8001 T; 1525 other;
 Query Match 31.2%; Score 708.6; DB 22; Length 32174;
 Best Local Similarity 99.4%; Pred. No. 5.1e-187;
 Matches 711; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1546 TTTCCTGTGGGCTGTGGGAGAGTGTGTGAGAAATCTCTCTCTTTCACAGCAGC 1605
 DB 26125 TCTTGAGGTGGCTGTGGGAGAGTGTGTGAGAAATCTCTCTCTTTCACAGCAGC 26184
 QY 1606 TCTTTTGGCCCAAGAGAGAAATGAAAAAGCAGGAGGTGAGATGATGCTTCACG 1665
 DB 26185 TCTTTTGGCCCAAGAGAGAAATGAAAAAGCAGGAGGTGAGATGATGCTTCACG 26244
 QY 1666 TGTGCTCTGTGCTGCCAGCAAGTCTTCAATTTGGGGCCAAAGGGGAACTTTTGTGAG 1725
 DB 26245 TGTGCTCTGTGCTGCCAGCAAGTCTTCAATTTGGGGCCAAAGGGGAACTTTTGTGAG 26304
 QY 1726 AAGGCGCTTGTGCTTTGTGACCCAGCTGGAATGACAGTGGCGGGATCTCAGCTCACCGCAA 1785
 DB 26305 AAGGCGCTTGTGCTTTGTGACCCAGCTGGAATGACAGTGGCGGGATCTCAGCTCACCGCAA 26364
 QY 1786 CCTCAGCTTCTGTGGGTTCAAGTATTTCTGCTCTCAGCTCTCCAAAGTGTGGGAATAC 1845
 DB 26365 CCTCAGCTTCTGTGGGTTCAAGTATTTCTGCTCTCAGCTCTCCAAAGTGTGGGAATAC 26424
 QY 1846 AAGCAGCGCCACCATGCGCCAGCTAAATTTTGTGATTTTCAATGATAAAGCGGATTTACACAG 1905
 DB 26425 AAGCAGCGCCACCATGCGCCAGCTAAATTTTGTGATTTTCAATGATAAAGCGGATTTACACAG 26484
 QY 1906 TTGGCAGGCTGTGTCTGAACTCTGACCGCAAGTATCACCGCTCGCTCCCAA 1965
 DB 26485 TTGGCAGGCTGTGTCTGAACTCTGACCGCAAGTATCACCGCTCGCTCCCAA 26544
 QY 1966 GTGCTGGGATTAAGGCGTGAAGCAGCAGCGTCCCGGCCAAAGGGGAACTTTGTGGAG 2025
 DB 26545 GTGCTGGGATTAAGGCGTGAAGCAGCAGCGTCCCGGCCAAAGGGGAACTTTGTGGAG 26604
 QY 2026 GAGCAGAGGGGCTCACATCTCCCTGTGATTTCCCAATGACATGCTTATCTCCCC 2085
 DB 26605 GAGCAGAGGGGCTCACATCTCCCTGTGATTTCCCAATGACATGCTTATCTCCCC 26664
 QY 2086 ATCTAGCCAGAACTATGTTTCTTCTGCAATTTACTATGATGTGATGTGCC 2145
 DB 26665 ATCTAGCCAGAACTATGTTTCTTCTGCAATTTACTATGATGTGATGTGCC 26724
 QY 2146 GCTACCAACACCCCGCCATGCGGGGGGTGGAAGGGGTGCAAGGCCCTGCTCACT 2205
 DB 26725 GCTACCAACACCCCGCCATGCGGGGGGTGGAAGGGGTGCAAGGCCCTGCTCACT 26784
 QY 2206 TTTTCTACCTGGAATGATTAAGATTAATCACTCTGTTGTTCAGTTTCA 2260
 DB 26785 TTTTCTACCTGGAATGATTAAGATTAATCACTCTGTTGTTCAGTTTCA 26839

Search completed: November 22, 2003, 02:18:40
 Job time : 605 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2003, 02:06:23 ; Search time 4740 Seconds
(without alignments)
11554.868 Million cell updates/sec

Title: US-09-938-803-26

Perfect score: 2273

Sequence: 1 99999tgaagcgtacgtt.....ttttcaaaaaaaaaaaaaa 2273

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estro:*
8: em_hic:*
9: gp_est1:*
10: gp_est2:*
11: gp_hic:*
12: gp_est3:*
13: gp_est4:*
14: gp_est5:*
15: em_estfun:*
16: em_estom:*
17: em_ges_hum:*
18: em_ges_inh:*
19: em_ges_pln:*
20: em_ges_vrt:*
21: em_ges_fun:*
22: em_ges_mam:*
23: em_ges_mus:*
24: em_ges_pro:*
25: em_ges_rtd:*
26: em_ges_phg:*
27: em_ges_vrl:*
28: gp_ges1:*
29: gp_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1036.2	45.6	1201	13	BX402797 BX402797
2	1002	44.1	1130	9	AL550083 AL550083
3	952.4	41.9	1754	11	AK083687 Mus muscu
4	930.4	40.9	1000	9	AL532090 AL532090

5	907.6	39.9	1149	13	BX339313
6	906.4	39.9	1053	12	BM557200
7	906.2	39.9	958	12	AL532089
8	873.4	38.4	1201	9	AL557774
9	864.8	38.0	1065	9	AL574064
10	856.2	37.7	1201	9	AL579737
11	855.2	37.6	1201	9	AL576170
12	843.2	37.1	1097	12	BM553135
13	837.6	36.8	976	13	BX431852
14	836.4	36.8	974	13	BQ917856
15	828.4	36.4	882	12	B1767794
16	822.6	36.2	887	10	BG753831
17	818.2	36.0	911	14	CA489207
18	814.4	35.8	929	10	BG740629
19	806.6	35.5	916	13	BUS41837
20	800.2	35.2	905	13	BUS56427
21	794.4	34.9	847	12	B1760522
22	793.2	34.9	897	9	AU149574
23	789.8	34.7	1201	9	AL553069
24	789.4	34.7	1252	10	BG684877
25	777.8	34.2	800	12	BG82157
26	776.6	34.2	967	10	BG696449
27	772.4	34.0	893	13	BUS53426
28	765.4	33.7	884	12	B1199710
29	763.6	33.6	827	13	BUS97281
30	761	33.5	994	13	B0675002
31	758.8	33.4	791	9	AU126035
32	757.8	33.3	815	9	AU124644
33	757.6	33.3	858	10	BG757592
34	751.8	33.1	1069	13	BQ067563
35	751.6	33.1	947	13	BUS55890
36	750.8	33.0	787	9	AU130894
37	749.8	33.0	830	9	AU133043
38	744.8	32.8	835	9	AU143336
39	738.2	32.5	1134	12	BG824361
40	735.6	32.4	798	9	AU131310
41	733	32.2	744	13	BX098372
42	721	31.7	744	10	BG758511
43	715.2	31.5	787	9	AU125489
44	711.6	31.3	816	10	BG695903
45	709.8	31.2	737	12	BG772149

ALIGNMENTS

RESULT 1
BX402797/c
LOCUS
DEFINITION BX402797 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
Homo sapiens CDNA clone CSDDU003YA18 3-PRIME, mRNA sequence.
ACCESSION BX402797
VERSION BX402797.1 GI:30622908
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by life technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6809.f for
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSIAJ0012E10NP1cluster=6809.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS1AJ001ZE10NP1.

FEATURES
source
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ003YA18"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 290 a 246 c 355 g 265 t 45 others

ORIGIN

Query Match 45 6%; Score 1036.2; DB 13; Length 1201;
Best Local Similarity 92.3%; Pred. No. 5.3e-16;
Matches 1096; Conservative 37; Mismatches 47; Indels 7; Gaps 5;

1031 CATGAAGTGTGGCTCTTTGAACAACCCCGCTTCACTGCACCATTCCTCATTTCA 1090
1181 CAACATGAGGTGTTGTTTAAACCAACCCCGCTTCACTGCACCATTCCTCATTTCA 1124
1091 CATCAACACCAACGCTGGTGGCCGCTTACATCTTCAACGACTCAAGTTCTTTGAAA 1150
1123 CATCAACACCAACGCTT-GGKGGCCGCTTACATCTTCAACGACTCAAG-TCCTTGAAA 1066
1151 TAAAGACTCTCTCAGGAGTCTCGTTGATTCCTGGCCCTGCGGAGGCGCTGCATCT 1210
1065 TAAAGACTCTCTCAGGAGTCTCGTTGATTCCTGGCCCTGCGGAGGCGCTGCATCT 1006
1211 AGGATACCTGCTGCTTCCAGATGGA-TTCCATATTGTTATTTGAAAGACAGAGCTG 1269
1005 AGGATACCTGCTGCTTCCAGATGGA-TTCCATATTGTTATTTGAAAGACAGAGCTG 946
1270 CCAAGCTCATTTCAAGAGAGCCCACTGACAGAGCTGGCCGCTATTCCTCTCCAGC 1329
945 CCAAGCTCATTTCAAGAGAGCCCACTGACAGAGCTGGCCGCTATTCCTCTCCAGC 886
1330 CCTTCTACTATTGGTGCAACAGACCATCCAGGCTTCAATGGGTTACTTCATGACAG 1389
885 CCTTCTACTATTGGTGCAACAGACCATCCAGGCTTCAATGGGTTACTTCATGACAG 826
1390 CCTTCTGCTCTTCAAGTGGAACAAAGGCTTAAGGTATTAATCCATCTATTTCCTTG 1449
825 CCTTCTGCTCTTCAAGTGGAACAAAGGCTTAAGGTATTAATCCATCTATTTCCTTG 766
1450 GCCCATCTTCTCTGAGGCTTCAATCTATTCCTTATTCATTCACAAAGCATGCTG 1509
765 GCCCATCTTCTCTGAGGCTTCAATCTATTCCTTATTCATTCACAAAGCATGCTG 706
1510 CAAAGAAAGAGATTAAAGAAATGAATCAATTCCTCTGAGGCTGAGGCGGAG 1569
705 CAAAGAAAGAGATTAAAGAAATGAATCAATTCCTCTGAGGCTGAGGCGGAG 646
1570 TGGTGAGAAACTACTGCTCTCCCTTTTCAACAGACTCTTTGCCCAAGAGAGATG 1629
645 TGGTGAGAAACTACTGCTCTCCCTTTTCAACAGACTCTTTGCCCAAGAGAGATG 586
1630 GAAAAAGCAGGAGGAGTGAAGATGATCTTCAAGCTGCTCTGCTGCGACCAAGTC 1689
585 GAAAAAGCAGGAGGAGTGAAGATGATCTTCAAGCTGCTCTGCTGCGACCAAGTC 526
1690 TTTATTTGGGGCAAAAGGGAATCTTTTGGAGAGGAGCTTGGCTTTGCAACCCAC 1749
525 TTTATTTGGGGCAAAAGGGAATCTTTTGGAGAGGAGCTTGGCTTTGCAACCCAC 466
1750 GCTGAATGAGTGGCGGAGTCTCAAGCTCAACGCACTCTCTCTGGTTCAAGTGA 1809
465 GCTGAATGAGTGGCGGAGTCTCAAGCTCAACGCACTCTCTCTGGTTCAAGTGA 406

1810 TTTTCTGCTTCAAGCTTCCCAAGTACAGGCAATACAGGCAAGCCCACTGCGCACTAA 1869
405 TTTTCTGCTTCAAGCTTCCCAAGTACAGGCAATACAGGCAAGCCCACTGCGCACTAA 346
1870 TTTTGTATTTTCAAGTACAGGCAATACAGGCAATACAGGCAATACAGGCAATACAGG 1929
345 TTTTGTATTTTCAAGTACAGGCAATACAGGCAATACAGGCAATACAGGCAATACAGG 286
1930 TCAAGCAAGTATCCACCCGCTTCCGCTTCCCAAGTCTGGATTACAGGCGTGAAGC 1989
285 TCAAGCAAGTATCCACCCGCTTCCGCTTCCCAAGTCTGGATTACAGGCGTGAAGC 226
1990 ACCGTCGCCGCGCCCAAGGGAATCTTTGTGGAGAGACCAAGGCGCTCAATCTCCCG 2049
225 ACCGTCGCCGCGCCCAAGGGAATCTTTGTGGAGAGACCAAGGCGCTCAATCTCCCG 166
2050 TCTGATTCGCCCATGACATTCCTTATCTCTCCCATCTAGCCAGAACTATTGTGT 2109
165 TCTGATTCGCCCATGACATTCCTTATCTCTCCCATCTAGCCAGAACTATTGTGT 106
2110 TTTTCTTCCCAATTACTATGATTTGTATGTGCGGCTTACAGCAACCCCGCATGAGG 2169
105 TTTTCTTCCCAATTACTATGATTTGTATGTGCGGCTTACAGCAACCCCGCATGAGG 46
2170 GAGTGAGAGGAGTGAAGGCGCTGCTGCTCACTTTTCTACCTT 2216
45 SSTGCGMCCGCGCCCAAGGCGCGCGCC-CTCCACATCAATCACTT 1

RESULT 2
AL550083 1130 bp mRNA linear EST 31-MAY-2003
LOCUS
DEFINITION
AL550083 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0DI040YC05 5-PRIME, mRNA sequence.
ACCESSION
VERSION
AL550083.2 GI:31271901
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
Li, W.B., Gruber, C., Jessee, J. and Poyles, D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 15, 2001 this sequence version replaced gi:12886703.
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6809.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DI040AB03QPL&cluster=6809.f. Contact:
Feng Liang Email: fliang@life.techn.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DI040AB03QPL.
FEATURES
source
1. .1130
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI040YC05"
/issue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 240 a 292 c 294 g 299 t 5 others

JOURNAL
REFERENCE
AUTHORS
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1754)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuta, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imclari, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, M., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Takami, D., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M., and Hayashizaki, Y.

TITLE
JOURNAL
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers

FEATURES

source

1..1754
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM.DB:D030066012"
/db_xref="taxon:10090"
/clone="D030066012"
/clone_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="9 days embryo"
<1..1125
/note="unnamed protein product; putative
unknown EST (GB|BF302972, evidence: BLASTN, 96%,
match=736)"
/codon_start=1
/protein_id="BAC38993.1"
/db_xref="GI:26350713"
/translation="VITLCPQMAVYLAGYYATGVDYDKMTMPHCVLTLKLGICI
DYDGGKDNLSLTSBOOKYAIRGVPSLLVAGSVPYGAFLVQPSMHHKYLVRGO
LIDIPKRNSTIPALKRLSLGLVYVGTLLSPHITDYLTDIDNRPFRCKYIM
LIMGRFVLKRYVTCWLVTEGVCLISGLNGFDENGSTVMDACANRWLFETPRFN
GTIASFNINMTAMVARYIFKRLKFLNGKLSQGLSLFLALMGLHSGVLIQFQMF
IYVKEKQVSLIRDSPLASLSTIALQFVYVQDTIMLFEGYSMTAFCLTMDKM
LKYRSIYFLGHVFFSLFLPIYIKHAWPRKXKKKE"
1731..1736
/note="putative"
polyA_site
1754
/note="putative"
polyA_site
1754
/note="putative"
BASE COUNT 437 a 477 c 383 g 457 t
ORIGIN

Query Match 41.9%; Score 952.4; DB 11; Length 1754;
Best Local Similarity 85.4%; Pred. No. 2.9e-14;
Matches 1090; Conservative 0; Mismatches 171; Indels 15; Gaps 2;

QY 417 GTCCTACTACCTTGGCTTCAGATGGCTACTCTTGCGTGGATATTACCTGCG 476
DB 1 GTTATTACTACCTTGGCTTCAGATGGCTACTCTTGCGGATATTACTACAGCC 60
QY 477 ACCGCACTACGATATCAAGTGAGCAATGCAATGTGTGCTTGAAGTGAAT 536
||||| ||||||| ||||||| ||||||| ||||| |||||||

DB 61 ACCGGTACTACGATATCAAGTGAGCAATGCAATGTGTGCTTGAAGTGAAT 120
QY 537 GCTTGGCTGTGACTACTCTTGAACGAGGGAAGATCAGATTCTGTCTGTGACAA 596
DB 121 GGGCTGTGATTGACTACTACGATGAGGCAAGACGGGAATCTCTGACTGTGACAA 180
QY 597 CAGAAATATGCAATACGTGTGCTTCCCTGCTGGAAGTGTGCTTCTCTTCTTC 656
DB 181 CAGAAATATGCAATACGTGTGCTTCCCTGCTGGAAGTGTGCTTCTCTTCTTC 240
QY 657 TATGGGCTTGTGTTAGGAGGCGCAAGTCTCAATGAACTACATGAAAGTGTGACG 716
DB 241 TATGAGCTTCTGTGTTAGGAGGCGCAAGTCTCAATGAACTACATGAAAGTGTG 300
QY 717 GAGAGCTATGACATACGAGAAAGATACCAACAGATATCTGCTCTCAAGCGC 776
DB 301 GAGAGCTATGACATACGAGAAAGATACCAACAGATATCTGCTCTCAAGCGC 360
QY 777 CTGAGCTGTGAGCTTTTCTACCTAGTGGGCTACACAGCTGAGCCCGCAATCAGAA 836
DB 361 CTGAGCTGTGAGCTTTTCTACCTAGTGGGCTACACAGCTGAGCCCGCAATCAGAC 420
QY 837 GACTATCTCTCACTGAGACTATGACCAACACCCCTTCTGCTCGCTGATGTACATG 896
DB 421 GACTATCTCTCACTGAGACTATGACCAACACCCCTTCTGCTCGCTGATGTACATG 480
QY 897 CTGATCTGGGCAAGTGTGCTGTGCAATATGTCACCTGTGCTGTGCTACAGAGA 956
DB 481 CTGATCTGGGCAAGTGTGCTGTGCAATATGTCACCTGTGCTGTGCTACAGAGA 540
QY 957 GTATGCAATTTGACGGGCTGTGAGCTTCAATGCTTGAAGAAAGGCAAGCAAGTGG 1016
DB 541 GTGTGCTGTGCTGTGAGGCTGTGAGCTTCAATGCTTGAAGAAAGGCAAGCAAGTGG 600
QY 1017 GATGCTGTGCAACATGAGAGTGTGCTCTTTGAAACCAACCCCGCTTCACTGGCACC 1076
DB 601 GATGCTGTGCAACATGAGAGTGTGCTCTTTGAAACCAACCCCGCTTCACTGGCACC 660
QY 1077 ATTGCTCTATTCAACATCAACCAACGCTGTGGTGGCGCTTACATCTTCAAGACATC 1136
DB 661 ATGCGCTCTTCAACATCAACCAACGCTGTGGTGGCGCTTACATCTTCAAGACATC 720
QY 1137 AAGTTCCTTGAAGAAATTAAGAACTCTCTCAGGCTCTGCTGCTATCTCTGAGCCCTCTGG 1196
DB 721 AAGTTCCTTGAAGAAATTAAGAACTCTCTCAGGCTCTGCTGCTATCTCTGAGCCCTCTGG 780
QY 1197 CACGGCTGTGCACTACGAGATACCTGTGCTCTTCAAGATGAATCTCTATGTTGTG 1256
DB 781 CATGGCTGTGCACTACGAGATACCTGTGCTCTTCAAGATGAATCTCTATGTTGTG 840
QY 1257 GAAAGACAGGCTGTGAGGCTCATTTCAAGAGACCCCACTTACAGCAAGTGTGCGCCGAT 1316
DB 841 GAAAGACAGGCTGTGAGGCTCATTTCAAGAGACCCCACTTACAGCAAGTGTGCGCCGAT 900
QY 1317 ACTGTCCTCAGACCCCTTCTACTATTGTGGAACAAGACATCACTGAGCTCTTCAATGGT 1376
DB 901 ACTGTCCTCAGACCCCTTCTACTATTGTGGAACAAGACATCACTGAGCTCTTCAATGGT 960
QY 1377 TACTCATGACTGCTCTTCTGCTCTTCACTGCTGGAACAATGAGCTTAAAGTGTAAATGC 1436
DB 961 TACTCATGACTGCTCTTCTGCTCTTCACTGCTGGAACAATGAGCTTAAAGTGTAAATGC 1020
QY 1437 ATCTATTTCTGTGCGCAACATCTTTCTGTGAGCCCTATATATTTGCTTATATTCAC 1496
DB 1021 ATCTATTTCTGTGCGCAACATCTTTCTGTGAGCCCTATATATTTGCTTATATTCAC 1080
QY 1497 AAAGCAATGTGCAAGAAAGAAAGAAAGTAAAGAGATGAATATC-----CATTTGCC 1551
DB 1081 AAAGCAATGTGCAAGAAAGAAAGAAAGTAAAGAGATGAATATC-----CATTTGCC 1140
QY 1552 TGTGCGCTGTGCGGAGCTGTGAGAAAGTAACTGCTGCTCTTCAAGCAAGCTCTTT 1611
DB 1141 TGTGCGCTGTGCGGAGCTGTGAGAAAGTAACTGCTGCTCTTCAAGCAAGCTCTTTCA 1200

QY 1612 GCCCAGAGCAGAGATGGAAAAAGCCAGAGAGTGAGATGATGCTTCCAGCTGTGCC 1671
DB 1201 CCCC-----AGCAGAGAGCCAGAGAGCTGGAAGTCAAGACAGACTCTCCAGCTGTGCC 1250
QY 1672 TCTGCTGCCAGCCAG 1687
DB 1251 TCTGCGACAGCCCAAG 1266

RESULT 4
AL532090 1000 bp mRNA linear EST 23-MAY-2003
LOCUS AL532090 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
DEFINITION CS0DM004YH02 5-PRIME, mRNA sequence.
ACCESSION AL532090
VERSION AL532090.2 GI:31069922
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1000)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 13, 2001 this sequence version replaced gi:12795583.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6809.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DM004YH02&cluster=6809.f. Contact :
Feng liang Email : fliang@life-tech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DM004YH02.1QPI.
Location/Qualifiers
1.1000
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM004YH02"
/issue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT 210 a 266 c 247 g 272 t 5 others
ORIGIN

Query Match 40.9%; Score 930.4; DB 9; Length 1000;
Best Local Similarity 98.5%; Pred. No. 1.4e-13;
Matches 938; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

QY 22 GCCCGCATTCGGGGGCGCGGAGCTGAGGGGGTCCCTGTGGGGCTCCCGAGTTAAGATG 81
DB 49 GTCGGGAATTCGGGGATCGAGAGTGGGGGTCCCTGTGGGGCTCCCGAGTTAAGATG 108
QY 82 GCTCTCAGCGGAGGGGAGCAGAGGAGCTGTGTGGCGCTGGCGGGGGTTCTGCACTGG 141
DB 109 GCTCTCAGCGGAGGGGAGCAGAGGAGCTGTGTGGCGCTGGCGGGGGTTCTGCACTGG 168
QY 142 GTTTCAGAGAGCTGAGCTTAACTTAAGATTGGGAGCGTCCCTGGGCGCGTCAAGACGGCGC 201
DB 169 GTTTCAGAGAGCTGAGCTTAACTTAAGATTGGGAGCGTCCCTGGGCGCGTCAAGACGGCGC 228
QY 202 TGGCGTAGTCATCTCCATCTTCTGGGTACCCCTTGGCTTTGTTTATCGGCAATTACC 261

DB 229 TGGCGTAGTCATCTCCATCTTCCGGGTACCCCTTGGCTTTGTTTATCGGCAATTACC 288
QY 262 TTTTCTAAGAGAGACCTACTCTATCCACCTCTTTCATACCTTTACAGGCGCTTCATTTG 321
DB 289 TTTTCTAAGAGAGACCTACTCTATCCACCTCTTTCATACCTTTACAGGCGCTTCATTTG 348
QY 322 CTTATTTTAACTTTGGAACAGAGCTCTACCACTCCCTGCTGTGATGTTGCTTCACTTCC 381
DB 349 CTTATTTTAACTTTGGAACAGAGCTCTACCACTCCCTGCTGTGATGTTGCTTCACTTCC 408
QY 382 TCATCTTTCGACTAATGAGGCGGACATCACTGCGCTCTGACTACCTTTTCTTCCAGA 441
DB 409 TCATCTTTCGACTAATGAGGCGGACATCACTGCGCTCTGACTACCTTTTCTTCCAGA 468
QY 442 TGGCTTACTCTTGGCTGTGATTAATTAACCTGCGGCACTAGATATCAAGTGA 501
DB 469 TGGCTTACTCTTGGCTGTGATTAATTAACCTGCGGCACTAGATATCAAGTGA 528
QY 502 CAATGCCAATTTGTTCTGACTTTGAAGCTGATGTTGGCTGTGACTTACTTTCAGG 561
DB 529 CAATGCCAATTTGTTCTGACTTTGAAGCTGATGTTGGCTGTGACTTACTTTCAGG 588
QY 562 GAGGAGAAAGATCAGAAATTCCTTGTCTCTGAGCAACAGAAATATGCAATAGCTGTTC 621
DB 589 GAGGAGAAAGATCAGAAATTCCTTGTCTCTGAGCAACAGAAATATGCAATAGCTGTTC 648
QY 622 CTTCCTCTCTGGAAGTTGCTGTCTTCTCTACTTCTATAGGCGCTTCTTGGAGGCCCC 681
DB 649 CTTCCTCTCTGGAAGTTGCTGTCTTCTCTACTTCTATAGGCGCTTCTTGGAGGCCCC 708
QY 682 AGTTCTCAATGAATCACTACATGAAGCTGTGAGGAGAGCTGATTAACATCCAGGA 741
DB 709 AGTTCTCAATGAATCACTACATGAAGCTGTGAGGAGAGCTGATTAACATCCAGGA 768
QY 742 AGATACCAAGACGATCATCTCTGCTCAAGCGCTGAGTCTGGGCGCTTTTCTACCTAG 801
DB 769 AGATACCAAGACGATCATCTCTGCTCAAGCGCTGAGTCTGGGCGCTTTTCTACCTAG 828
QY 802 TGGGTATACACTGTCTGCGCCCGGACATCAAGAGATATCTCTCACTGAAGACTATG 861
DB 829 TGGGTATACACTGTCTGCGCCCGGACATCAAGAGATATCTCTCACTGAAGACTATG 888
QY 862 ACAACCAACCCCTTCTGCTGCGCTGATGATGATGAGGCGAAGTTTGGCTGT 921
DB 889 ACAACCAACCCCTTCTGCTGCGCTGATGATGATGAGGCGAAGTTTGGCTGT 948
QY 922 ACAATATGTACCTGTGTGCTGTGTCAGAGAGAGTATGATTTTGAACGG 973
DB 949 ACAATATGTACCTGTGTGCTGTGTCAGAGAGAGTATGATTTTGAACGG 1000

RESULT 5
BX339313 1149 bp mRNA linear EST 02-MAY-2003
LOCUS BX339313 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0DI070Y016 5-PRIME, mRNA sequence.
ACCESSION BX339313
VERSION BX339313.1 GI:30335787
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1149)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 6809.f. For more information about this cluster, see

<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0D1070DE080P1a&cluster=6809.f>. Contact : Feng Liang Email : liang@lifetech.com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0D1070DE080P1.

FEATURES

Source

1..1149

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0D1070XJ16"

/issue_type="PLACENTA COT 25-NORMALIZED"

/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT

240 a 285 c 285 g 298 t 41 others

ORIGIN

Query Match 39.9%; Score 907.6; DB 13; Length 1149;
Best Local Similarity 97.9%; Pred. No. 3.9e-13;
Matches 924; Conservative 5; Mismatches 14; Indels 1; Gaps 1;

23 CCCGATTTGGGGCGCGGAGTGGGGGCTCTGTGGGGCTCCCGAGTTAAGATGCG 82
53 CCGGAATTCGGGGATCGCGAGCTGGGGGCTCCCTGGGGCTCCCGAGTTAAGATGCG 112
83 GTCCTCAGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 142
113 GTCCTCAGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 172
143 TTTCAGAGAGCTGAGGCTTACAGTGTGCGAGCTCCCTGGGGCGCTCAGAACAGGCGCT 202
173 TTTCAGAGAGCTGAGGCTTACAGTGTGCGAGCTCCCTGGGGCGCTCAGAACAGGCGCT 232
203 GGGGCTGATCATCTCCATCTTCTCTGGGTTACCCCTTGTCTTTTATGCGATTACCT 262
233 GGGGCTGATCATCTCCATCTTCTCTGGGTTACCCCTTGTCTTTTATGCGATTACCT 292
263 TTTCAGAGAGAGAGCTACCTCACTACAGCTCTCCATCTTACAGGAGGCTCCATTCG 322
293 TTTCAGAGAGAGAGCTACCTCACTACAGCTCTCCATCTTACAGGAGGCTCCATTCG 352
323 TTATTTAATTTGGAACAGCTCTACAGCTCCCTGCTGTGATTTGCTTCACTTCT 382
353 TTATTTAATTTGGAACAGCTCTACAGCTCCCTGCTGTGATTTGCTTCACTTCT 412
383 CATCTCTGACATTAATGAGGCGGAGCACTACCTGCGCTCTCACTACCTTTTCTCCAGAT 442
413 CATCTCTGACATTAATGAGGCGGAGCACTACCTGCGCTCTCACTACCTTTTCTCCAGAT 472
443 GGGCTACCTTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 502
473 GGGCTACCTTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 532
503 AATGACACATTTGTTCTGACTTTGAAGCTGATTTGGCTTTGACTTGAAGCTTGAAG 562
533 AATGACACATTTGTTCTGACTTTGAAGCTGATTTGGCTTTGACTTGAAGCTTGAAG 592
563 AAGGAAAGATCAGAAATTCCTTGTCTGAGCAACAGAAATATGCAATAGTGTGTTCC 622
593 AAGGAAAGATCAGAAATTCCTTGTCTGAGCAACAGAAATATGCAATAGTGTGTTCC 652
623 TTCCCTGCTGGAAGTTGCTGTTCTCACTTCTAATGAGGAGGAGGAGGAGGAGGAG 682
653 TTCCCTGCTGGAAGTTGCTGTTCTCACTTCTAATGAGGAGGAGGAGGAGGAGGAG 712
683 GTTCTCAATGATCCTACATGAGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 742
713 GTTCTCAATGATCCTACATGAGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 772

QY 743 GATACCAACAGACATCATTTCTGCTCTCAAGGCGCTGAGCTGGGCTTTTCTACTACT 802
DB 773 GATACCAACAGACATCATTTCTGCTCTCAAGGCGCTGAGCTGGGCTTTTCTACTACT 832
QY 803 GGGCTACACAGCTGCTCAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 862
DB 833 GGGCTACACAGCTGCTCAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 892
QY 863 CAACACCCCTTCTGTTCCGCTGAGTATGCTGATCTGGGCAAGTTTGTCTGTA 922
DB 893 CAACACCCCTTCTGTTCCGCTGAGTATGCTGATCTGGGCAAGTTTGTCTGTA 952
QY 923 CAATATGCTACCTGTTGCTGCTGCTACAGAGAGGATATGATTT 966
DB 953 CAATATGCTACCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 995

RESULT 6

BM557200

LOCUS

DEFINITION

AGENCOURT 6579051 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:546619

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LCM1967 row: e column: 20
High quality sequence stop: 677.
Location/Qualifiers

FEATURES

source

1..1053

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:546619"

/tissue_type="amelanotic melanoma, cell line"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH MGC 41"

/note="Organ: skin; Vector: pORF7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC library"

BASE COUNT

230 a 313 c 237 g 267 t 6 others

ORIGIN

Query Match 39.9%; Score 906.4; DB 12; Length 1053;
Best Local Similarity 95.4%; Pred. No. 4.4e-13;
Matches 962; Conservative 0; Mismatches 42; Indels 4; Gaps 3;

QY 1092 ATCAACACCAACGCTGGGCTGCGGCTACATCTTCAACGACTCAAGTTCTTGAAT 1151
DB 1 ATCAACACCAACGCTGGGCTGCGGCTACATCTTCAACGACTCAAGTTCTTGAAT 60

OY	1152	AAAGAACTCTCTCAGAGGTCCTCGTGGTATTCCTGGACCTGGACAGGCTGCATCA	121
Db	61	AAAGAACTCTCTCAGAGGTCCTCGTGGTATTCCTGGACCTGGACAGGCTGCATCA	120
OY	1212	GGATACCTGGTCCTGCCTCAGATGGAATTCCTCATTTATATGTGGAAAGACAGCTGCC	127
Db	121	GGATACCTGGTCCTGCCTCAGATGGAATTCCTCATTTATATGTGGAAAGACAGCTGCC	180
OY	1272	AGGCTCATTCAGAGAGGCCCACTCGTAGCAAGCTGGCCGCACTATCTGCTCCAGCCC	1331
Db	181	AGGCTCATTCAGAGAGGCCCACTCGTAGCAAGCTGGCCGCACTATCTGCTCCAGCCC	240
OY	1332	TTCTACTATTTGGTGCACAAGACATCCAGCTGGGCTTCATGGGTATCTCCATGACTGCC	139
Db	241	TTCTACTATTTGGTGCACAAGACATCCAGCTGGGCTTCATGGGTATCTCCATGACTGCC	300
OY	1392	TTCTGCTCTTTCACGTGGGACAAATAGGCTTAAGGTATTAATTCATCTATTCCTTGAC	145
Db	301	TTCTGCTCTTTCACGTGGGACAAATAGGCTTAAGGTATTAATTCATCTATTCCTTGAC	360
OY	1452	CACATCTTCTTCTGAGCCTACTATTCATATTCCTTATATTCACAAAGCAATGGTGC	1511
Db	361	CACATCTTCTTCTGAGCCTACTATTCATATTCCTTATATTCACAAAGCAATGGTGC	420
OY	1512	AGGAAGAAGAGTTAAAGAGATGGAATTAATTCATTCCTCGTGGGCTGTGGGGAGCTG	1571
Db	421	AGGAAGAAGAGTTAAAGAGATGGAATTAATTCATTCCTCGTGGGCTGTGGGGAGCTG	480
OY	1572	GTGCAGAAACTACTGCTCTCCCTTTTCACAGCACTCTTTGGCCGAGAGCAGAAATGSA	1631
Db	481	GTGCAGAAACTACTGCTCTCCCTTTTCACAGCACTCTTTGGCCGAGAGCAGAAATGSA	540
OY	1632	AAAGCCAGAGAGGTGGAAGATGATGCTTCCAGCTGTGTCTGTGCACGCCAAGTCTT	1691
Db	541	AAAGCCAGAGAGGTGGAAGATGATGATGCTTCCAGCTGTGTCTGTGCACGCCAAGTCTT	600
OY	1692	CATTGGGGGGCAAAAGGGGAAACTTTTTTTGGAGAAAGGCTTGTGTCACCCACAGC	1751
Db	601	CATTGGGGGGCAAAAGGGGAAACTTTTTTTGGAGAAAGGCTTGTGTCACCCACAGC	660
OY	1752	TGGAATGCAGTGGCCGGATCTCAGCTCAACCGAACCTCCACCTCTGGGTTCAATGATTT	1811
Db	661	TGGAATGCAGTGGCCGGATCTCAGCTCAACCGAACCTCCACCTCTGTGTTCAATGATTT	720
OY	1812	TTCTGCTCAGCCTCCCAAGTACGTTGGGAATACAGGACGCGACATGCCAGCTAATTT	1871
Db	721	TTCTGCTCAGCCTCCCAAGTACGTTGGGAATACANGACGCGACATGCCAGCTAATTT	780
OY	1872	TTTGTATTTTTCAGTAGAAACGGGATTTTACCAACGTTGGCCAGGCTGTCTGAAATCTCTG	1931
Db	781	TTTGTATTTTTCAGTAGAAACGGGATTTTACCAACGTTGGCCAGGCTGTCTGAAATCTCTG	840
OY	1932	ACCGCAATGATTCACACGCGCTCCGCTCCCAAGTGTGGGATTACA-GGGGTAGAGCA	1992
Db	841	ACCGCGCAGTATTCACACGCGCTCCGCTCCCAAGTGTGGGATTACAAGGCGTATAGCCA	900
OY	1991	CCGTGCCCCGGCCCAAAAGGGGAAACTCTTG--TGAGAGAGCAGAGGGCTCATCT--CC	2047
Db	901	CCGTGCCCCGGCCCAAAAGGGGAAACTCTTGTGGAGNACAAAGGGGCTCATCTCC	960
OY	2048	CCTCTGATTTCCCCCATGCACATTTGCTTATCTCTCCCATCTAGCCAG	2095
Db	961	CCTCTGATTTCCCCCATGCACATTTGCTTATCTCTCTCCCATCTAG	1008

RESULT 7	AL532089/c	958 bp	mRNA	linear	EST 23-MAY-2003
LOCUS	AL532089				
DEFINITION	AL532089 Homo sapiens FETAL LAYER Homo sapiens CDNA clone				
ACCESSION	CSDDM004Y02.3-PRIME.				
VERSION	AL532089				
KEYWORDS	AL532089.2	GI:31069921			
	EST.				

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 958)
L1, W. B., Gruber, C., Jassae, J. and Polyes, D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 13, 2001 this sequence version replaced gi:12795582.

cgf-bin/cluster;cgf1seq=CSDDM004DD01NP1&cluster=6809.f. Contact:
Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope, sequence ID: CSDDM004DD01NP1.

FEATURES	Location/Qualifiers
source	1. .958

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODM004YH02"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_1lb="Homo sapiens FETAL LIVER"
/note="Organ: Liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-Oligo(dt) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT      240 a.    229 c    212 g    210 t      7 others

```

BASE COUNT	240 a	229 c	272 g	210 t	7 others
ORIGIN					
Query Match	39.9%	Score 906.2	DB 9	Length 958	
Best Local Similarity	98.1%	Pred. No. 4.8e-13			
Matches 942	Conservative 4	Mismatches 11	Indels 3	Gaps 3	

Qy	1245	ATTGTATTTGGAAGACAGAGCTGCCAGGCTCATTCAGAGAGGCCACCCCTGAGCAAG	1300
Dp	958	AATTGATATGTGGAAAGACAGAGCTGCCAGGCTCATTCAGAGAGGCCACCCCTGAGCAAG	899
Qy	1305	CTGGCCGCGCATTACTGTCTCTCCAGCCCTTCTACTATTTGTGTGCAACAGACCATCCACTGG	1364
Dp	898	CTGGCCGCGCATTACTGTCTCTCCAGCCCTTCTACTATTT - GAGCAACAGACCATCCACTGG	840
Qy	1365	CTCTTCATGGGTTACTGCATGACTGCTTCTGCGCTTCCAGGTGGGACAAATGGCTTAAAG	1422
Dp	839	CTCTTCATGGG - TACTCCATGACTGCTTCTGCGCTTCCAGGTGGGACAAATGGCTTAAAG	781
Qy	1425	GTCATATAATGCATCTATTTTCTTGGGCACATCTTCTCTGAGCGTCTACTATTCAATTG	1488
Dp	780	GTCATATAATGCATCTATTTCTTGGGCACATCTTCTCTGAGCGTCTACTATTCAATTG	721
Qy	1485	CTTATATATTCACAAAGCAGATGCTGCCAAGAAAGAGATTAAAGAGATGGAATATCC	1544
Dp	720	CTTATATATTCACAAAGCAGATGCTGCCAAGAAAGAGATTAAAGAGATGGAATATCC	661
Qy	1545	ATTTCGCCGTGGGACTGTGCGGGACTGGTGTGAGAACTACTCGTCTCCTTTTCACAGA	1600
Dp	660	ATTTCGCCGTGGGACTGTGCGGGACTGGTGTGAGAACTACTCGTCTCCTTTTCACAGA	601
Qy	1605	CTCCTTTTGCCCCAGAGCAGAGATGAAAGAACCCAGGAGGTGAGAGATCGATGCTTCCAG	1664
Dp	600	CTCCTTTTGCCCCAGAGCAGAGATGAAAGAACCCAGGAGGTGAGAGATCGATGCTTCCAG	541
Qy	1665	CTGTGCTCTGTCTGCCACCAAGTCTTCAATTTGGGGCCAAAGGGGAAAC - TTTT TTTTGG	1722
Dp	540	CTGTGCTCTGTCTGCCACCAAGTCTTCAATTTGGGGCCAAAGGGGAAAC TTTT TTTTGG	481

QY 1724 AGAAGCGTCTTCTTGTACCCAGCGTGAATGCAATGGGGGATCTACGCTACCC 1783
 DB 480 AGAAGCGTCTTCTTGTACCCAGCGTGAATGCAATGGGGGATCTACGCTACCC 421
 QY 1784 AACCTCACCCTCTGGGTTCAAGTATTTCTGCTCAGCCCTCCAGTACGCTGGGAT 1843
 DB 420 AACCTCACCCTCTGGGTTCAAGTATTTCTGCTCAGCCCTCCAGTACGCTGGGAT 361
 QY 1844 ACAGGACGCCACCATGAGCCAGCTAATTTTGTATTTGATAGAAACGGGATTTGACA 1903
 DB 360 ACAGGACGCCACCATGAGCCAGCTAATTTTGTATTTGATAGAAACGGGATTTGACA 301
 QY 1904 CGTTGGCCAGCGTGTCTGAACTCTGACCCGCAAGTATTCACCCGCTCGCTCCCA 1963
 DB 300 CGTTGGCCAGCGTGTCTGAACTCTGACCCGCAAGTATTCACCCGCTCGCTCCCA 241
 QY 1964 AAGTGTGAGATTACAGGGGTGAGCAAGTGTGCGGCGCCCAAGGGGAACTGTGGG 2023
 DB 240 AAGTGTGAGATTACAGGGGTGAGCAAGTGTGCGGCGCCCAAGGGGAACTGTGGG 181
 QY 2024 AGAGACAGAGGGGCTCACATCTCCCTCTGATTCGCCCATGACATGCTTATCTCTCC 2083
 DB 180 AGAGACAGAGGGGCTCACATCTCCCTCTGATTCGCCCATGACATGCTTATCTCTCC 121
 QY 2084 CCATCTAGCCAGGAATCTATGTTGTTTCTTCTGCAATTAATGATGATGATG 2143
 DB 120 CCATCTAGCCAGGAATCTATGTTGTTTCTTCTGCAATTAATGATGATGATG 61
 QY 2144 CGGCTACCAACACCCCGCATGGGGGGGAGAGAGGGGCTGCTGCTGCTCA 2203
 DB 60 CGGCTACCAACACCCCGCATGGGGGGGAGAGAGGGGCTGCTGCTGCTCA 1

RESULT 8
 AL557774 1201 bp mRNA linear EST 31-MAY-2003
 LOCUS AL557774 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
 DEFINITION Homo sapiens cDNA clone CS0DJ003YA18 5-PRIME, mRNA sequence.
 ACCESSION AL557774
 VERSION AL557774.2 GI:31279574
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1201)
 Full-length cDNA libraries and normalization
 Unpublished
 On Feb 15, 2001 this sequence version replaced gi:12901710.
 CONTACT: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: sequef@genoscope.cns.fr, Web: www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 6809.f For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0DJ003BA09Q1&cluster=6809.f. Contact :
 Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID: CS0DJ003BA09Q1.
 Location/Qualifiers
 1. 1201
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DJ003YA18"
 /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
 /cell_line="JURKAT"
 /clone_1b="Homo sapiens T CELLS (JURKAT CELL LINE) COT
 10-NORMALIZED"

FEATURES

source

/note="1st strand cDNA was primed with a NotI-oligo (dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."
 BASE COUNT 229 a 315 c 282 g 304 t 71 others

Query Match 38.4%; Score 873.4; DB 9; Length 1201;
 Best Local Similarity 91.0%; Pred. No. 2.2e-12;
 Matches 956; Conservative 32; Mismatches 50; Indels 12; Gaps 6;

QY 29 TTCCGGGGGCGCGGCTGGGGGGGCTCCCTGATGGGGCTCCGGAGTTAAGATGCGCGCTC 88
 DB 13 TCCCGGAGTCCCGAGCTGAGGAGGCTCCCTGATGGGGCTCCGGAGTTAAGATGCGCTC 72
 QY 89 ACCGAGGGGGAGCAGAGGGGAGCTGTGTGGCGCTGGCGGGGGTTCTGCACTCGGGTTTCCA 148
 DB 73 ACCGAGGGGGAGCAGAGGGGAGCTGTGTGGCGCTGGCGGGGGTTCTGCACTCGGGTTTCCA 132
 QY 149 GGAAGCTGAGCTTAAACAAGTTGGCGAGTCTCTGGGGCGGCTGAGAACAGGGCGCTGGGCT 208
 DB 133 GGAGCTGAGCTTAAACAAGTTGGCGAGTCTCTGGGGCGGCTGAGAACAGGGCGCTGGGCT 192
 QY 209 GATCATCTCCATCTTCCCTGGGTTACCCCTTGTGTTTATCGGCAATTAATCTTTCTTA 268
 DB 193 GATCATCTCCATCTTCCCTGGGTTACCCCTTGTGTTTATCGGCAATTAATCTTTCTTA 252
 QY 269 CAAGAGACCTACCTATCCATCCATCTTCAATCAAGGCTCTCAATGCTTAAT 328
 DB 253 CAAGAGACCTACCTATCCATCCATCTTCAATCAAGGCTCTCAATGCTTAAT 312
 QY 329 TAACCTTTGGAAACCAAGCTTACCACTCCCTGCTGTGATTTGCTTCACTTCTCATCT 388
 DB 313 TAACCTTTGGAAACCAAGCTTACCACTCCCTGCTGTGATTTGCTTCACTTCTCATCT 372
 QY 389 TCGACATAATGGGCGGACCATCACTGCGGCTCTCACTACCTTTTGGCTTCAGATGGCTTA 448
 DB 373 TCGACATAATGGGCGGACCATCACTGCGGCTCTCACTACCTTTTGGCTTCAGATGGCTTA 432
 QY 449 CCTTCTGGCTGATATCTATTTACATGCGCACCGGCACTACGATATCAAGTGAACAATGCC 508
 DB 433 CCTTCTGGCTGATATCTATTTACATGCGCACCGGCACTACGATATCAAGTGAACAATGCC 492
 QY 509 ACATTTGTCTGATCTTTGAAGCTGATGTTGGCTGTGACTACTTGAAGAGGAGGA 568
 DB 493 ACATTTGTCTGATCTTTGAAGCTGATGTTGGCTGTGACTACTTGAAGAGGAGGA 552
 QY 569 ACATGAGAAATCTTGTCTCTGAGCAACAGAAATATGCGCATACGTTGCTTCCCT 628
 DB 553 ACATGAGAAATCTTGTCTCTGAGCAACAGAAATATGCGCATACGTTGCTTCCCT 612
 QY 629 GCTGGAAGTTGCTGTTTCTCTACTTCTATGAGGAGCTTCTGTAAGGAGCCCACTTCTC 688
 DB 613 GCTGGAAGTTGCTGTTTCTCTACTTCTATGAGGAGCTTCTGTAAGGAGCCCACTTCTC 672
 QY 689 AATGATATCTATCAATGAAGCTGTGAGGAGAGCTGATGACATACCGAGAAATATAC 748
 DB 673 AATGATATCTATCAATGAAGCTGTGAGGAGAGCTGATGACATACCGAGAAATATAC 732
 QY 749 AAACAGCATATCTCTGCTCTGAGCAACAGAAATATGCGCATACGTTGCTTCCCT 808
 DB 733 AAACAGCATATCTCTGCTCTGAGCAACAGAAATATGCGCATACGTTGCTTCCCT 792
 QY 809 CACACTGTCAAGCCCGCACATGACAGAAAGTATCTCTCACTGAAGATATGACACCA 868
 DB 793 CACACTGTCAAGCCCGCACATGACAGAAAGTATCTCTCACTGAAGATATGACACCA 851
 QY 869 CCCCTTTGCTGCTGCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 928
 DB 852 CCCCTTTGCTGCTGCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 907
 QY 929 TGTCACTGTTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 988

Db 908 TATGKACACGTTGGYKCMCAGAGA-A-KATGSMITTTGACGGC--SGGGYTCATATG 964
Qy 989 CTTTGAAGAAAAGGCAAGCAAGTGGATGCTGTGCCAATGAGTGTGCTCTT 1048
Db 965 YTT-GAAGAAARGGCAAGGCA--KTGGWGCCTTGTGCMHAGAGAKTKGKCTTTTG 1020
Qy 1049 TGAACAAACCCCGCTTCACTGACCAT 1078
Db 1021 AAMMACCCCGCTCCCGGCGCTTGTCTTT 1050

RESULT 9
AL574064/c 1065 bp mRNA linear EST 31-MAY-2003
LOCUS AL574064 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0D1040YC05 3-PRIME, mRNA sequence.
ACCESSION AL574064
VERSION AL574064.2 GI:31295399
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1065)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 16, 2001 this sequence version replaced gi:12933905.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6809.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1040AB03NP1;cluster=6809.f. Contact :
Feng Liang Email : fliang@life.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0D1040AB03NP1.
Location/Qualifiers
1..1065
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1040YC05"
/cissue_type="PLACENTA COT 25-NORMALIZED"
/note="11b-Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 259 a 205 c 302 g 228 t
ORIGIN

Query Match 38.0%; Score 864.8; DB 9: Length 1065;
Best Local Similarity 91.8%; Pred No. 3.7e-12;
Matches 896; Conservative 28; Mismatches 45; Indels 3; Gaps 3;

Qy 1240 TCCATATTGTAATTTGGAAGACAGGCTGCCAGCTCATTCAGAGCCCACTCTGA 1299
Db 973 TTCCCATTTKATTTGTGAAAGAGGCTRCMAAATCTATTCAAGAGACCCCACTCTGA 914
Qy 1300 GCAAGCTGGCCGCTTACTGTCTCCAGCCCTTCTACTATTGTGTGCAACAGACATTC 1359
Db 913 GCAABBTGGCCGCTACTGTCTCCAGCCCTTCTACTATTGTGTGCAACAGACATTC 854
Qy 1360 ACTGCTCTTCACTGAGTACTCATGAGCTGCTTCTGCTTCTGAGGCAATGAC 1419
Db 853 ACTGCTCTTCTGCTGAGTACTCATGAGCTGCTTCTGCTTCTGAGGCAATGAC 795
Qy 1420 TTAAGTGTATAATTCATTAATTTCTTGGACACATCTTCTCTGAGCCCTACTATTC 1479

Db 794 TTAAGTGTATAATTCATTAATTTCTTGGACACATCTTCTCTGAGCCCTACTATTC 735
Qy 1480 TATTGCTTTATATTCACAAAGCATGTGCGCAAGAAAGAGTTAAAGAGTGAAT 1539
Db 734 TATTGCTTTATATTCACAAAGCATGTGCGCAAGAAAGAGTTAAAGAGTGAAT 675
Qy 1540 AATCATTTTCCCTGTGAGCTGTGCGGAGACTGTGAGAAATCTGCTCTCTCTTTCA 1599
Db 674 AATCATTTTCCCTGTGAGCTGTGCGGAGACTGTGAGAAATCTGCTCTCTCTTTCA 615
Qy 1600 CAGCATCTCTTGTGCGGAGACTGTGAGAAATCTGAGAAATCTGAGTGAATGATGCT 1659
Db 614 CAGCATCTCTTGTGCGGAGACTGTGAGAAATCTGAGAAATCTGAGTGAATGATGCT 555
Qy 1660 TCCAGCTGTGCTGTGCGGAGACTGTGAGAAATCTGAGAAATCTGAGTGAATGATGCT 1719
Db 554 TCCAGCTGTGCTGTGCGGAGACTGTGAGAAATCTGAGAAATCTGAGTGAATGATGCT 495
Qy 1720 TTGAGAAAGGCGTCTTGTGTCACCCAGCTGGAATGAGTGGCGGATCTGAGTCA 1779
Db 494 TTGAGAAAGGCGTCTTGTGTCACCCAGCTGGAATGAGTGGCGGATCTGAGTCA 435
Qy 1780 CCGCACTCTCACTCTCTGAGTGAATGATGATGATGATGATGATGATGATGATGATG 1839
Db 434 CCGCACTCTCACTCTCTGAGTGAATGATGATGATGATGATGATGATGATGATGATG 375
Qy 1840 GAATACAGGACCGGACCATGCGGATGATGATGATGATGATGATGATGATGATGATG 1899
Db 374 GAATACAGGACCGGACCATGCGGATGATGATGATGATGATGATGATGATGATGATG 315
Qy 1900 ACCAGTGTGCGGAGCTGTGTCACCTCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1959
Db 314 ACCAGTGTGCGGAGCTGTGTCACCTCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 255
Qy 1960 CCGCAAGTGTGAGTATACAGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 2019
Db 254 CCGCAAGTGTGAGTATACAGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 195
Qy 2020 TGGAGAGGAGGAGGAGGCTCATCTCCCTGATGCTCCCGACATGACATGCTTATCT 2079
Db 194 TGGAGAGGAGGAGGAGGCTCATCTCCCTGATGCTCCCGACATGACATGCTTATCT 135
Qy 2080 CTCCTCATGAGCAGGATCTATGTTTCTTCTGCAATTTACTATGATGTTGTA 2139
Db 134 CTCCTCATGAGCAGGATCTATGTTTCTTCTGCAATTTACTATGATGTTGTA 75
Qy 2140 TGTGCGCTTACACACACCCCGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2199
Db 74 TGTGCGCTTACACACACCCCGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 17
Qy 2200 TCCACTTTTCTACT 2215
Db 16 TCCACTTTTCTACT 1

RESULT 10
AL579737/c 1201 bp mRNA linear EST 01-JUN-2003
LOCUS AL579737 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0D1003VA18 3-PRIME, mRNA sequence.
ACCESSION AL579737
VERSION AL579737.2 GI:31318017
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 16, 2001 this sequence version replaced gi:12945073.
Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: sequef@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6809.f. For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DJ003BA09NP1&cluster=6809.f. Contact :
Feng Liang Email: fliang@life.techn.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DJ003BA09NP1.

FEATURES

source

```
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ003YA18"
/cell_type="T CELLS (URKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="URKAT"
/clone_11b="Homo sapiens T CELLS (URKAT CELL LINE) COT
10-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
```

BASE COUNT

258 a 247 c 334 g 250 t 112 others

ORIGIN

Query Match 37.7%; Score 856.2; DB 9; Length 1201;
Best Local Similarity 90.2%; Pred. No. 5.2e-12;
Matches 881; Conservative 50; Mismatches 42; Indels 4; Gaps 3;

```
QY 1216 ACCGTGCTGCTCCAGATGAATTCCTATTGTTATGTGGAAGACAGCTGCCAGGC 1275
DB 976 MNCGGCTGKGYTCYRGA--GGAATCCTCCATGTTGTTGTAACACDCKTKCCRC 919
QY 1276 TCATTGAAGAGCCCGCCAGAGAGCTGCCCGCCATTAATGCTCCACCCCTTCT 1335
DB 918 TCATTCARRKATACCCCACTAGAGAGKGCGCCCATWAKTCTCCASCTCYT 859
QY 1336 ACTATTGTGTCAGACAGACCATCAGCTGCTCTTCATAGGTTACTCCAGTCTCTCT 1395
DB 858 ACTATTGTGTCAGACAGMCAVCATGCTGCTTCATAGTTCACTCCATGACGCTTCT 799
QY 1396 GCTCTTTCACGTGGCAAAATGCTTAAGGTATTAATCATCTATTTCTCTGG--CCAC 1454
DB 798 GCCTCTTTCACGTGGCAAAATGCTTAAGGTATTAATCATCTATTTCTCTGGSCAC 739
QY 1455 ATCTCTCTGCGCTACTATTCATATGCTTATATTCACAAAGCATGTCGCAAG 1514
DB 738 ATCTCTCTGCGCTACTATTCATATGCTTATATTCACAAAGCATGTCGCAAG 679
QY 1515 AAAGAGAAGTTAAAGATGATATATCCATTTCCCTGGTGGCTGTGGGAGTGGTG 1574
DB 678 AAAGAGAAGTTAAAGATGATATATCCATTTCCCTGGTGGCTGTGGGAGTGGTG 619
QY 1575 CAGAAACTACTGCTCTCTCTTTCACAGCACTCTTTCGCCCGACAGAGAGAAATGAAAA 1634
DB 618 CAGAAACTACTGCTCTCTCTTTCACAGCACTCTTTCGCCCGACAGAGAGAAATGAAAA 559
QY 1635 GCCAGAGAGGTGAAGATGATGATCTTCAGCTGCTGCTGCGCAGCAAGCTTCTCAT 1694
DB 558 GCCAGAGAGGTGAAGATGATGATCTTCAGCTGCTGCTGCGCAGCAAGCTTCTCAT 499
QY 1695 TTGGGGCCAAAGGGGAACTTTTGTGAGAGAGGGCTTGTGCTTGTCAACCCAGCTGG 1754
DB 498 TTGGGGCCAAAGGGGAACTTTTGTGAGAGAGGGCTTGTGCTTGTGMAACCAAGCTGG 439
QY 1755 AATGCACTGGCGGAGATCTACGTCACCGCACTCCACTCTCTGGGTTCAAGTATTTTC 1814
DB 438 AATGCACTGGCGGAGATCTACGTCACCGCACTCCACTCTCTGGGTTCAAGTATTTTC 379
QY 1815 CTGCTCAGCTCCCAAGTATGCTGGGAATACAGGACGCAACCATGCCCAGCTAATTTT 1874
```

```
DB 378 CTGCTCAGCTCTCCAGATGCTGGAAACAGGACGCCACACATGCCCAGCAATTTT 319
QY 1875 GTATTTTTCAGTAAAGGGATTTTTCACAGCTTTGGCCAGCTGTCTCTCACTCTGACC 1934
DB 318 GTATTTTTCAGTAAAGGGATTTTTCACAGCTTTGGCCAGCTGTCTCTCACTCTGACC 259
QY 1935 GCAAGTATTCACCCGCTCTCCCTCCCAAGTCTGGATTAACAGGCTGACCACTGT 1994
DB 258 GCAAGTATTCACCCGCTCTCCCTCCCAAGTCTGGATTAACAGGCTGACCACTGT 199
QY 1995 GCCCGGCCCAAGGGGAAACTTTGAGGAGAGGAGGAGGCTACATCTCCCTCTGA 2054
DB 198 GCCCGGCCCAAGGGGAAACTTTGAGGAGAGGAGGAGGATCACTCTCCCTCTGCA 139
QY 2055 TTCCCGCATGACATTTGCTTCTCTCCCATCTAGCCAGATCTATTTGTTTTCT 2114
DB 138 TTCCCGCATGACATTTGCTTCTCTCCCGAAMMAGCAGGAATCTATTTGTTTTCT 79
QY 2115 TCTGCCAATTCTATGATTTGTTATGTCGCTTACACACACCCCTCATGGGGGGTG 2174
DB 78 TCTGCCAATTCTATGATTTGTTATGTCGCTTACACACACCCCTCATGGGGGGTG 20
QY 2175 GAGAGGGGTCGCAAGGCC 2191
DB 19 GAGAGGGGTCGCAAGGCC 3
```

RESULT 11

AL576170/c 1201 bp mRNA linear EST 01-JUN-2003

LOCUS AL576170 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA

DEFINITION clone CS0DI072YO19 3-PRIME, mRNA sequence.

ACCESSION AL576170.2 GI:31314460

VERSION AL576170.2

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1201)

AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished

COMMENT On Feb 16, 2001 this sequence version replaced gi:12938048.

Contact: Genoscope
Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France
Email: sequef@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6809.f. For
more information about this cluster, see

http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DI072AH10NP1&cluster=6809.f. Contact :

Feng Liang Email: fliang@life.techn.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DI072AH10NP1.

Location/Qualifiers

1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI072YO19"
/issue_type="PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 284 a 261 c 322 g 266 t 68 others

ORIGIN

Query Match 37.6%; Score 855.2; DB 9; Length 1201;

Beet Local Similarity 95.0%; Pred. No. 5.5e-12;
Matches 896; Conservative 13; Mismatches 30; Indels 4; Gaps 3;

```

QY 1303 AGCTGGCCGCACTTACTGCTCCAGCCCTTACTATTTGGTGCACAGACCATTCCT 1362
DB 943 AGMAYSTGSCCATPAATGCTCCCTCAACCTTACTATTTGGTGCACAGACCATTCCT 884
QY 1363 GGCCTTCATGSGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1422
DB 883 GGCCTTCATGSGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 824
QY 1423 AGGTATTAATTCATCTATTTT-CCTTGGCCATCTTCTCTGAGCTTACTATTCATA 1481
DB 823 AGGTATTAATTCATCTATTTTCCCTTGGCCATCTTCTCTGAGCTTACTATTCATA 764
QY 1482 TTGCTTTATTCACAAAGCATGTGTCACAAAGAAAGAAAGTAAAGATGTGAATA 1541
DB 763 TTGCTTTATTCACAAAGCATGTGTCACAAAGAAAGAAAGTAAAGATGTGAATA 704
QY 1542 TCATTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1601
DB 703 TCATTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 644
QY 1602 GCATCTCTTTTGGCCAGAGAGAAATGAGAAAGCCAGAGGTGAGATGATGCTTC 1661
DB 643 GCATCTCTTTTGGCCAGAGAGAAATGAGAAAGCCAGAGGTGAGATGATGCTTC 584
QY 1662 CAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1721
DB 583 CAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 524
QY 1722 GGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1781
DB 523 GGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 464
QY 1782 GCAACCTCCACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1841
DB 463 GCAACCTCCACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 404
QY 1842 ATACAGAGCAGCCAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1901
DB 403 ATACAGAGCAGCCAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 344
QY 1902 CAGCTTGGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1961
DB 343 CAGCTTGGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 284
QY 1962 CAAAGTGTGAGATTACAGCGTGAAGCCAGCGTGGCCGCAAGAGGGAATCTGCT 2021
DB 283 CAAAGTGTGAGATTACAGCGTGAAGCCAGCGTGGCCGCAAGAGGGAATCTG 224
QY 2022 GGAGAGAGCAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2081
DB 223 GGAGAGAGCAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 164
QY 2082 CCCCATCTAGCAGGAATCTATTTGTTTCTTCTGCAATTTACTATGATGTGATG 2141
DB 163 CCCCATCTAGCAGGAATCTATTTGTTTCTTCTGCAATTTACTATGATGTGATG 104
QY 2142 TGCCGCTACACCA--CCCCCCCCATGGGGGGGTGAGAGGGGTGCAAGGCTGCTG 2199
DB 103 TGCCGCTACACCAAGAGGCGCCCCCAATGGGGGGGTGAGAGGGGTGCAAGGCTGCTG 44
QY 2200 TCCAC-TTTTCTTACTGTTGAAGTGTATTAATAATCACTT 2241
DB 43 TCCAC-TTTTCTTACTGTTGAAGTGTATTAATAATCACTT 1

```

RESULT 12
BM553135 1097 bp mRNA linear EST 20-FEB-2002
LOCUS BM553135
DEFINITION AGENCOURT_6572558 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5467062
5', mRNA sequence.

```

ACCESSION BM553135
VERSION BM553135.1 GI:18791600
KEYWORDS EST
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS 1 (bases 1 to 1097)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LCM1968 row: h column: 07
High quality sequence scop: 666.
Location/Qualifiers
1..1097
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5467062"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1ib="NIH_MGC_41"
/note="Organ: skin; Vector: pOB7; Site 1: XhoI; Site 2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 244 a 315 c 254 g 275 t 9 others
ORIGIN
Query Match 37.1%; Score 843.2; DB 12; Length 1097;
Beet Local Similarity 94.5%; Pred. No. 1.1e-11;
Matches 937; Conservative 0; Mismatches 41; Indels 14; Gaps 6;
QY 1092 ATCAACACCAACGCTGGGTGGCCGCTACATCTTCAACGACTCAAGTTCCTTGAAT 1151
DB 1 ATCAACACCAACGCTGGGTGGCCGCTACATCTTCAACGACTCAAGTTCCTTGAAT 60
QY 1152 AAAGAACTCTCAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1211
DB 61 AAAGAACTCTCAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
QY 1212 GGATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1271
DB 121 GGATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
QY 1272 AGGCTCATTCAGAGAGCCCAACCTTGAGCAAGCTGGCCGCCATTACTGCTCCAGGCC 1331
DB 181 AGGCTCATTCAGAGAGCCCAACCTTGAGCAAGCTGGCCGCCATTACTGCTCCAGGCC 240
QY 1332 TTCTACTATTTGGTGCACAGACCATTCAGCTGCTCTTCAATGGTTACTCCATGACTGCC 1391
DB 241 TTCTACTATTTGGTGCACAGACCATTCAGCTGCTCTTCAATGGTTACTCCATGACTGCC 300
QY 1392 TTCTGCTCTTCACGCTGGGCAAAATGGCTTAAGGTGTAATAATCCATCTATTTCTTGCC 1451
DB 301 TTCTGCTCTTCACGCTGGGCAAAATGGCTTAAGGTGTAATAATCCATCTATTTCTTGCC 360
QY 1452 CACATCTTCTTCCTGAGCCTACTATTCATATTCCTTATTCACAAAGCAATGGTGCCA 1511
DB 361 CACATCTTCTTCCTGAGCCTACTATTCATATTCCTTATTCACAAAGCAATGGTGCCA 420

```

QY 1512 AGGAAAGAGATTTAAAGAGATGATATATCCATTTCCCTGGTGGCTGTGGGACTG 1571
 DB 421 AGGAAAGAGATTTAAAGAGATGATATATCCATTTCCCTGGTGGCTGTGGGACTG 480
 QY 1572 GTGACAAATCTACTGCTCCCTTTTCAAGCACTCTTGGCCAGAGAGAGATGGA 1631
 DB 481 GTGACAAATCTACTGCTCCCTTTTCAAGCACTCTTGGCCAGAGAGAGATGGA 540
 QY 1632 AAAGCAGGAGGTGGAAGATGATGATCTTCCAGCTGTGCTGTGCGCAGCAAGCTT 1691
 DB 541 AAAGCAGGAGGTGGAAGATGATGATCTTCCAGCTGTGCTGTGCGCAGCAAGCTT 600
 QY 1692 CATTTGGGGCCAAAGGGGAACTTTTGTGAGAAAGGCTCTGCTTGTACCCAGC 1751
 DB 601 CATTTGGGGCCAAAGGGGAACTTTTGTGAGAAAGGCTCTGCTTGTACCCAGC 660
 QY 1752 TGGATGACAGTGGGGGATCTCAGCTCACCGCAACCTCCACCTGCTGGGTTCAAGTAT 1811
 DB 661 TGGATGACAGTGGGGGATCTCAGCTCACCGCAACCTCCACCTGCTGGGTTCAAGTAT 720
 QY 1812 TTCTGCTCTCAGCTCCCAAGTATGCTGGGATACAGGCAAGCCAGCAGTGGCTAT 1871
 DB 721 TTCTGCTCTCAGCTCCCAAGTATGCTGGGATACAGGCAAGCCAGCAGTGGCTAT 779
 QY 1872 TTTGTATTTTTCAGTAAACGGGATTTTCAACAGTGGTGGCAGGC-TGCTCTCAACTCT 1930
 DB 780 TTTGTATTTTTCAGTAAACGGGATTTTCAACAGTGGTGGCAGGC-TGCTCTCAACTCT 839
 QY 1931 GACCGGAGTATGCAAGCC--GGCTCGGCTCCCAAGTATGCT--GGGATTAAGAGGCT--G 1985
 DB 840 GACCGGAGTATGCAAGCC--GGCTCGGCTCCCAAGTATGCT--GGGATTAAGAGGCT--G 1985
 QY 1986 AGCCACCGTGGCCCGCCCAAGGAGAACTTTGTG-----GAGAGACAGAGGGGCT 2038
 DB 900 AGCCACCGTGGCCCGCCCAAGGAGAACTTTGTG-----GAGAGACAGAGGGGCT 2038
 QY 2039 CACATCTCCCTCTGATTTCCCATGACATTT 2070
 DB 960 CATCTTCCCTCTGATTTCCCATGACATTT 991

RESULT 13
 LOCUS BX431852 876 bp mRNA linear EST 15-MAY-2003
 DEFINITION BX431852 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
 CS0DM004YH02 5-PRIME, mRNA sequence.
 ACCESSION BX431852
 VERSION BX431852.1 GI:30779040
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1. (bases 1 to 876)
 Full-length cDNA libraries and normalization
 Unpublished
 COMMENT
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segre@genoscope.cns.fr, Web: www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 6809.f. For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS08AG05ZE03 CS06165 1acluster=6809.f.
 Contact: Peng Liang Email: fliang@life.com URL:
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID: CS08AG05ZE03_CS06165_1.
 Location/Qualifiers
 1. 876
 /organism="Homo sapiens"

/mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DM004YH02"
 /issue_type="FETAL LIVER"
 /dev_stage="fetal"
 /clone_lib="Homo sapiens FETAL LIVER"
 /note="Organ: liver; Vector: pcwvSPORT_6; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-strand cDNA was digested with Not I and
 cloned into the Not I and EcoRV sites of the pcwvSPORT 6
 vector. Library was not normalized."
 BASE COUNT 207 a 237 c 201 g 227 t 4 others
 ORIGIN
 Query Match 36.8%; Score 837.6; DB 13; Length 876;
 Best Local Similarity 99.0%; Pred. No. 1.7e-11;
 Matches 862; Conservative 0; Mismatches 7; Indels 2; Gaps 2;
 QY 842 TCTCTCACTGAAGACTATGACACCAACCCCTTGTGCTCCGCTGATGATGCTGAT 901
 DB 3 TCTCTCACTGAAGACTATGACACCAACCCCTTGTGCTCCGCTGATGATGCTGAT 62
 QY 902 CTGGGCAAGTTGTGCTGTACAAATATGTCACTGTTGGTGTGTCACAGAGAGTATG 961
 DB 63 CTGGGCAAGTTGTGCTGTACAAATATGTCACTGTTGGTGTGTCACAGAGAGTATG 122
 QY 962 CATTTGACGGGCTTGGGCTTCAATGCTTTGAAGAAAGGCAAGCAAGTGGGATGC 1021
 DB 123 CATTTGACGGGCTTGGGCTTCAATGCTTTGAAGAAAGGCAAGCAAGTGGGATGC 182
 QY 1022 CTGTGCAATGAAGGTGTGCTCTTGAACAAACCCCGCTTCACTGGGCAACCTTC 1081
 DB 183 CTGTGCAATGAAGGTGTGCTCTTGAACAAACCCCGCTTCACTGGGCAACCTTC 242
 QY 1082 CTCAATCAATCAACCAACGAGCTGGTGGCCGCTCACTCTTCAACGACTCAAGTT 1141
 DB 243 CTCAATCAATCAACCAACGAGCTGGTGGCCGCTCACTCTTCAACGACTCAAGTT 302
 QY 1142 CTTTGAATTAAGAACTTCTGAGGGTCTCTGTTGCTATTCCTGCGCTTGGGACGG 1201
 DB 303 CTTTGAATTAAGAACTTCTGAGGGTCTCTGTTGCTATTCCTGCGCTTGGGACGG 362
 QY 1202 CCTGCACTGAGATACCTGCTGCTCCAGATGGAATCCATTTGATTTGGAAG 1261
 DB 363 CCTGCACTGAGATACCTGCTGCTCCAGATGGAATCCATTTGATTTGGAAG 422
 QY 1262 ACAGGCTGCAAGCTCATTAAGAGAGCCCAACCTGAGCAAGCTGGCCGCTTACTG 1321
 DB 423 ACAGGCTGCAAGCTCATTAAGAGAGCCCAACCTGAGCAAGCTGGCCGCTTACTG 482
 QY 1322 CTTCCAGCCCTTCTACTATTTGGTGCACAGACATCCATGCTCTTCAATGGCTTACT 1381
 DB 483 CTTCCAGCCCTTCTACTATTTGGTGCACAGACATCCATGCTCTTCAATGGCTTACT 542
 QY 1382 CATGACTGCTTCTGCTCTTCAAGTGGGACAAAGGCTTAAGGTATTAATCCATCTA 1441
 DB 543 CATGACTGCTTCTGCTCTTCAAGTGGGACAAAGGCTTAAGGTATTAATCCATCTA 602
 QY 1442 TTTCTTGGCCACATCTTCTCTAGCCCTACTATTAATGCTTATTAATCAACAAG 1501
 DB 603 TTTCTTGGCCACATCTTCTCTAGCCCTACTATTAATGCTTATTAATCAACAAG 662
 QY 1502 AATGTGCAAGAGAAAGAGATTAAGAGATGGAATATCCATTTCCCTGTGGCTG 1561
 DB 663 AATGTGCAAGAGAAAGAGATTAAGAGATGGAATATCCATTTCCCTGTGGCTG 722
 QY 1562 TGGGGAGCTGGTGCAGAACTACTGCTCCCTTTTCAAGAGACTCTTGGCCGAGAG 1621
 DB 723 TGGGGAGCTGGTGCAGAACTACTGCTCCCTTTTCAAGAGACTCTTGGCCGAGAG 782
 QY 1622 AGAGATGAGAAAGCCAGGAGAGT--GGAAGATGATGCTTCCAGCTGTGCTGTGCGC 1680
 DB 783 AGAGATGAGAAAGCCAGGAGAGT--GGAAGATGATGCTTCCAGCTGTGCTGTGCGC 842

QY 1681 AGCCAACTCTTCATTTGGGGCCAAAGGGGAA 1711
 DB 843 AGCCAAAGTC-TCAATTTGGGCCAAAGGGGAA 872

RESULT 14
 LOCUS BQ917856 974 bp mRNA linear EST 20-AUG-2002
 DEFINITION AGENCOURT 8801447 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6311255
 5' mRNA sequence.
 ACCESSION BQ917856
 VERSION BQ917856.1 GI:22332554
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 TITLE NIH-MGC http://mgi.nci.nih.gov/.
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Susan L. Sullivan, PhD.
 cDNA Library Preparation: Resgen, Invitrogen Corp.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Plate: LLM13734 row: n column: 24
 High quality sequence stop: 755.
 Location/Qualifiers
 1..974
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6311255"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC_129"
 /note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1.cdb; Site: 1; ScoreV; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: this is a NIH MGC Library."
 is a NIH MGC Library."

BASE COUNT 199 a 260 c 248 g 267 t
 ORIGIN

Query Match 36.8%; Score 836.4; DB 13; Length 974;
 Best Local Similarity 97.2%; Pred. No. 1.7e-11;
 Matches 883; Conservative 0; Mismatches 21; Indels 4; Gaps 3;

QY 33 GGGCGCGGAGCTGGGGGGGCTCTGTGGGCTCCCGAGTTAAGATGCGTCTCAGG 92
 DB 1 GGTACGCGAGGCTGGGTGGGCAATTCGCGACAGGGGAGATTAGATGCGTCTCAGG 60

QY 93 GAGGGGAGCAGAGGAGCTGTGGCGCTGGGGGGGTTCTCAGTCCGGTTTCCAGAG 152
 DB 61 GAGGGGAGCAGAGGAGCTGTGGCGCTGGGGGGGTTCTCAGTCCGGTTTCCAGAG 120

QY 153 CTGAGCCTTAAACAATTGGCGAGTCCCTGGGCGCTGAGACAGGCGCTGGGCGTATC 212
 DB 121 CTGAGCCTTAAACAATTGGCGAGTCCCTGGGCGCTGAGACAGGCGCTGGGCGTATC 180

QY 213 ATCTCATCTTCTGGGTTACCCCTTTGTTTATCGGCATTACTCTTTTCAAG 272
 DB 181 ATCTCATCTTCTGGGTTACCCCTTTGTTTATCGGCATTACTCTTTTCAAG 240

QY 273 GAGACTACCTCATCACTCTTCACTTACCTTACAGGCTCTCAATGCTATTATAC 332
 DB 241 GAGACTACCTCATCACTCTTCACTTACCTTACAGGCTCTCAATGCTATTATAC 300

QY 333 TTGGAAACAGAGCTCTACCACTCCCTGCTGTATTTGTTCACTTCTCATCTTGA 392
 DB 301 TTGGAAACAGAGCTCTACCACTCCCTGCTGTATTTGTTCACTTCTCATCTTGA 360

QY 393 CTATGGGCGGACCATCATCTGCGCTCTCACTACCTTTTGGTTCCAGATGGCTTACT 452
 DB 361 CTATGGGCGGACCATCATCTGCGCTCTCACTACCTTTTGGTTCCAGATGGCTTACT 420

QY 453 CTGGGTGATCTATTAACCTGCGACCGGCAACTGATCAATCAATGACAAATGCAAT 512
 DB 421 CTGGGTGATCTATTAACCTGCGACCGGCAACTGATCAATCAATGACAAATGCAAT 480

QY 513 TGTGTTGACTTTGAACTGATTTGTTGGCTGTGACTACTTTGACGAGGAGAAAT 572
 DB 481 TGTGTTGACTTTGAACTGATTTGTTGGCTGTGACTACTTTGACGAGGAGAAAT 540

QY 573 CAGAAATTCCTTGTCTCTGAGCAACAAGAAATATGCAATAGTGTTCTTCTCTGCTG 632
 DB 541 CAGAAATTCCTTGTCTCTGAGCAACAAGAAATATGCAATAGTGTTCTTCTCTGCTG 600

QY 633 GAAATGCTGCTTCT 692
 DB 601 GAAATGCTGCTTCT 660

QY 693 AATCACTACATGAAGCTGTGAGAGAGAGAGCTGATTGACATACAGAAAGATACCAAC 752
 DB 661 AATCACTACATGAAGCTGTGAGAGAGAGAGCTGATTGACATACAGAAAGATACCAAC 720

QY 753 AGCATATTCCTGCTCTCAAGGCGCTGAGTCTGGGCTTTTCTTACTGATGGGCTACAA 812
 DB 721 AGCATATTCCTGCTCTCAAGGCGCTGAGTCTGGGCTTTTCTTACTGATGGGCTACAA 780

QY 813 CTGCTACAGCCCGCAATACAGAGAGATATCTCTCTCACTGAAAGATATAGACA-CCACC 871
 DB 781 CTGCTACAGCCCGCAATACAGAGAGATATCTCTCTCACTGAAAGATATAGACA-CCACC 840

QY 872 CTTCGTGTTCCGCTGATGATCAT-GCTGATCTGGG--CAAGTTTGCTGTACAATA 928
 DB 841 CTTCGTGTTCCGCTGATGATCATGATGCTGATGCTGGGGCCAAAGTTTGCTGTACAATA 900

QY 929 TGTCACT 936
 DB 901 TGGCACCT 908

RESULT 15
 B1767794 882 bp mRNA linear EST 25-SEP-2001
 LOCUS 603060787F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5210300 5',
 DEFINITION mRNA sequence.
 ACCESSION B1767794
 VERSION B1767794.1 GI:15759372
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 JOURNAL NIH-MGC http://mgi.nci.nih.gov/.
 COMMENT Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Plate: LLM1527 row: m column: 21
 High quality sequence stop: 840.
 Location/Qualifiers

FEATURES

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0D1040AB030P1.
Location/Qualifiers

FEATURES
source

1. 1130
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1040Y05"
/issue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 240 a 292 c 294 g 299 t 5 others

ORIGIN

Query Match 42.3%; Score 961; DB 9; Length 1130;
Best Local Similarity 100.0%; Pred. No. 1.5e-155;
Matches 961; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

47 GGGGGGTCCTGTTGGGCTCCGAGTTAAGTGGCTCCTCAGCGAGGGGAGCAAGG 106
124 GGGGGGTCCTGTTGGGCTCCGAGTTAAGTGGCTCCTCAGCGAGGGGAGCAAGG 183
107 GACTGTGTGGCCCTGGCGGGGTTCTGCAGTGGGTTCCAGAGAGCTGAGCCTTAACA 166
184 GACTGTGTGGCCCTGGCGGGGTTCTGCAGTGGGTTCCAGAGAGCTGAGCCTTAACA 243
167 GTTGGCGAGCTCCCTGGGCGCTCAGAACAGGCGCTGATCATCTCCATCTTCT 226
244 GTTGGCGAGCTCCCTGGGCGCTCAGAACAGGCGCTGATCATCTCCATCTTCT 303
227 GGGTTACCCCTTGTCTTTTATGCGATTACCTTTTACAGAGAGCCTACCTCAT 286
304 GGGTTACCCCTTGTCTTTTATGCGATTACCTTTTACAGAGAGCCTACCTCAT 363
287 CCACCTCTTCCATACCTTTACAGGCTCTCAATTTGCTTATTTAAGTGGAAACAGCT 346
364 CCACCTCTTCCATACCTTTACAGGCTCTCAATTTGCTTATTTAAGTGGAAACAGCT 423
347 CTACCACTCCCTGCTGT 406
424 CTACCACTCCCTGCTGT 483
407 CATCACTGCGCTCTCACTACCTTTTGTCCAGATGGCTTCTGCTGAGATCA 466
484 CATCACTGCGCTCTCACTACCTTTTGTCCAGATGGCTTCTGCTGAGATCA 543
467 TTACACTGCGACCGGCACTACGATATCAAGTGAAGCAATGCCATTTGTGTGTGT 526
544 TTACACTGCGACCGGCACTACGATATCAAGTGAAGCAATGCCATTTGTGTGTGT 603
527 GAAGCTGATGTTGT 586
604 GAAGCTGATGTTGT 663
587 CTCTGAGCAACAGAAATATGCCATACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 646
664 CTCTGAGCAACAGAAATATGCCATACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 723
647 CTCTGATCTTATGTGGGCTTTCTTTGTGAGGCGCCAGTTCTCAATGATCATACATGA 706
724 CTCTGATCTTATGTGGGCTTTCTTTGTGAGGCGCCAGTTCTCAATGATCATACATGA 783
707 GCTGTGACAGGAGAGCTGATTTGACATACAGGAAATATCCAAAGCAGTCATCTCTGC 766
784 GCTGTGACAGGAGAGCTGATTTGACATACAGGAAATATCCAAAGCAGTCATCTCTGC 843
767 TCTCAAGCGCTGAGTGTGGCCCTTTTCTACCTAGTGGGCTACACACTGTCTAGCCCCCA 826
844 TCTCAAGCGCTGAGTGTGGCCCTTTTCTACCTAGTGGGCTACACACTGTCTAGCCCCCA 903

QY 827 CATCAAGAGACTATCTCTCACTGAGACTATGACAACCAACCCCTTGTGTCGGCTG 886
DB 904 CATCAAGAGACTATCTCTCACTGAGACTATGACAACCAACCCCTTGTGTCGGCTG 963
QY 887 CATGTACATGCTGATCTGGGCGCAAGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 946
DB 964 CATGTACATGCTGATCTGGGCGCAAGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1023
QY 947 CACAGAGAGATGATGATTTTGAACGGGCTGGCTTCAATGGCTTTGAAGAAAGGCA 1006
DB 1024 CACAGAGAGATGATGATTTTGAACGGGCTGGCTTCAATGGCTTTGAAGAAAGGCA 1083
QY 1007 G 1007
DB 1084 G 1084

RESULT 2

AL532090

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Feb 13, 2001 this sequence version replaced gi:12795583.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6809.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DM004D010P1&cluster=6809.f. Contact :
Feng Liang Email: fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DM004D010P1.
Location/Qualifiers

FEATURES

source

1. 1000
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM004YH02"
/issue_type="FETAL LIVER"
/dev_stage="fetal"
/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo (dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

BASE COUNT 210 a 266 c 247 g 272 t 5 others

ORIGIN

Query Match 38.8%; Score 881; DB 9; Length 1000;
Best Local Similarity 100.0%; Pred. No. 7.5e-142;
Matches 881; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 GAGGGGACGAGGAGCTGTGTGGCGCTGGCGGGGTTCTGAGTCGGGTTCCAGAG 152
DB 120 GAGGGGACGAGGAGCTGTGTGGCGCTGGCGGGGTTCTGAGTCGGGTTCCAGAG 179
QY 153 CTGAGCTTAAAGAGTTGGGAGCTCCCTGGGCGCTGAGAACAGGCGCTGCGCTGATC 212


```

Db      180 CTGAGCTTAAAGATTGGGAGCTCCCTGGGGCGTCAAGACAGCGCTGCGCTGATC 239
QY      213 ATCTCATCTTCTGGGTTACCCCTTGTGTTTATCGGATTACCTTTCTACAG 272
Db      240 ATCTCATCTTCTGGGTTACCCCTTGTGTTTATCGGATTACCTTTCTACAG 299
QY      273 GAGACCTACCTTCATGACCTCTTTCATACCTTTACAGGCTCTCAATTGCTTATTTTAA 332
Db      300 GAGACCTACCTTCATGACCTCTTTCATACCTTTACAGGCTCTCAATTGCTTATTTTAA 359
QY      333 TTGGAACACAGCTCAACCTCCCTGCTGTATGTTGTGTTAGTTCTCATCTTTCGA 392
Db      360 TTGGAACACAGCTCAACCTCCCTGCTGTATGTTGTGTTAGTTCTCATCTTTCGA 419
QY      393 CTATATGGGCGGACCATCATCTGCGCTCTCATACCTTTTGTCTTCAGATGGCTACCT 452
Db      420 CTATATGGGCGGACCATCATCTGCGCTCTCATACCTTTTGTCTTCAGATGGCTACCT 479
QY      453 CTGGCTGATCTATTACACTGCGCACCGCACTACAGATATCAAGTGAACAATGCCACAT 512
Db      480 CTGGCTGATCTATTACACTGCGCACCGCACTACAGATATCAAGTGAACAATGCCACAT 539
QY      513 TGTGTTCTACCTTTGAAGCTGATTTGGTTGGCTGTGTTGACTTCTTACCGAGGAAAGAT 572
Db      540 TGTGTTCTACCTTTGAAGCTGATTTGGTTGGCTGTGTTGACTTCTTACCGAGGAAAGAT 599
QY      573 CAGAAATTCCTTCTCTGAGCAAGAAATATGCAATGCTGTGTTCTTCCCTGCTG 632
Db      600 CAGAAATTCCTTCTCTGAGCAAGAAATATGCAATGCTGTGTTCTTCCCTGCTG 659
QY      633 GAAGTTGCTGTTTCTCTCACTTCTATGGGCGCTTCTGTGAGGGCCGCACTTCTCAAT 692
Db      660 GAAGTTGCTGTTTCTCTCACTTCTATGGGCGCTTCTGTGAGGGCCGCACTTCTCAAT 719
QY      693 AATCATATATGAAGCTGTGCGAGGAGAGCTATTGACATACACAGAAAGATACCAAC 752
Db      720 AATCATATATGAAGCTGTGCGAGGAGAGCTATTGACATACACAGAAAGATACCAAC 779
QY      753 AGCATATCTCTGCTCTCAAGCGCTGAGTCTGGGCTTTTCTACCTAGTGGGCTACACA 812
Db      780 AGCATATCTCTGCTCTCAAGCGCTGAGTCTGGGCTTTTCTACCTAGTGGGCTACACA 839
QY      813 CTGCTCAGCCCCCAGATCAAGAAAGCTATCTCTCAGTGAAGACTATGACAAACACCC 872
Db      840 CTGCTCAGCCCCCAGATCAAGAAAGCTATCTCTCAGTGAAGACTATGACAAACACCC 899
QY      873 TTCTGTTCCGCTGATGTAACATGCTGATCTGGGGCAAGTTTGTCTGTACAATATGTC 932
Db      900 TTCTGTTCCGCTGATGTAACATGCTGATCTGGGGCAAGTTTGTCTGTACAATATGTC 959
QY      933 ACCCTGTGGCTGGTCAAGAGAGATGATTTTGAAGGG 973
Db      960 ACCCTGTGGCTGGTCAAGAGAGATGATTTTGAAGGG 1000

RESULT 3
BX431852 876 bp mRNA linear EST 15-MAY-2003
LOCUS     BX431852 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
DEFINITION
CS0DM004YH02 5-PRIME, mRNA sequence.
ACCESSION BX431852
VERSION   BX431852
KEYWORDS  BX431852.1 GI:30779040
SOURCE    EST.
ORGANISM  Homo sapiens (human)
           Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 876)
AUTHORS   Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished
COMMENT    Contact: Genoscope
           Genoscope - Centre National de Sequencage

```

```

BP 191 91006 EVRY cedex - France
Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6809.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAG065ZE03_CS06165_1&cluster=6809.f.
Contact : Feng Liang Email : fliang@life.techn.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0BAG065ZE03_CS06165_1.
FEATURES
source
1..876
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM004YH02"
/issue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT 207 a 237 c 201 g 227 t 4 others
ORIGIN
Query Match 34.5%; Score 784; DB 13; Length 876;
Best Local Similarity 100.0%; Pred. No. 2.9e-125;
Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
857 CTATAGCAACACCCCTTCTGGTCCGCTGATGTAACATGCTGATCGGGCAAGTTGT 916
Db 18 CTATAGCAACACCCCTTCTGGTCCGCTGATGTAACATGCTGATCGGGCAAGTTGT 77
QY 917 GCTGTACAATATGTCACCTGTTGGCTGATCAGAGAGATGATCTTTGACGGGCT 976
Db 78 GCTGTACAATATGTCACCTGTTGGCTGATCAGAGAGATGATCTTTGACGGGCT 137
QY 977 GGGCTTCAATGCTTTGAAAGAAAGGCAAGGCAAGTGGATGCTGTGCCAATGAA 1036
Db 138 GGGCTTCAATGCTTTGAAAGAAAGGCAAGGCAAGTGGATGCTGTGCCAATGAA 197
QY 1037 GGTGTGGCTCTTGAACAAACCCCGCTTCACTGGACCATTTGCTCAATCAATCA 1096
Db 198 GGTGTGGCTCTTGAACAAACCCCGCTTCACTGGACCATTTGCTCAATCAATCA 257
QY 1097 CACCAAGCCTGGTGGCCCGCTACATCTTCAAGAGACTCAAGTTCTTGAATAAAGA 1156
Db 258 CACCAAGCCTGGTGGCCCGCTACATCTTCAAGAGACTCAAGTTCTTGAATAAAGA 317
QY 1157 ACTCTCTAAGGCTCTCTGTTGCTATTTCTTGCCCTCTGGCAGCGCTGCACTAGA 1216
Db 318 ACTCTCTAAGGCTCTCTGTTGCTATTTCTTGCCCTCTGGCAGCGCTGCACTAGA 377
QY 1217 CCTGTGCTCTCCAGATGGAATTCCTCATGTTATTTGGAAGAAGCGCTGCCAGGT 1276
Db 378 CCTGTGCTCTCCAGATGGAATTCCTCATGTTATTTGGAAGAAGCGCTGCCAGGT 437
QY 1277 CATTCAGAAGAGCCCACTGAGCAAGCTGGCCGCAATTAATGCTCTCCAGCCCTTCTA 1336
Db 438 CATTCAGAAGAGCCCACTGAGCAAGCTGGCCGCAATTAATGCTCTCCAGCCCTTCTA 497
QY 1337 CTATTTGTTGCAACAGACCATTCACCTGCTCTTCAATGGGTTACTCATGACTGCTTCTG 1396
Db 498 CTATTTGTTGCAACAGACCATTCACCTGCTCTTCAATGGGTTACTCATGACTGCTTCTG 557
QY 1397 CCTCTTCAAGGAGCAATGGCTTAAGTGTATTAATCAATTAATTTCTTGGCAGCAT 1456
Db 558 CCTCTTCAAGGAGCAATGGCTTAAGTGTATTAATCAATTAATTTCTTGGCAGCAT 617
QY 1457 CTCTTCTGAGCCCTACTATTCATTTGCTTATATTCACAAGAGATGTCACAGGAA 1516

```

Db 618 CTCTTCTGAGGCTACTATTCATTGCTTATTCACAAAGCAATGGTGCAGGAA 677

QY 1517 AGAAGAGTTAAAGAAATGGAATATTCATTTCTCTGGTGCGCTGTGGGAGCTGTGCA 1576

Db 678 AGAAGAGTTAAAGAAATGGAATATTCATTTCTCTGGTGCGCTGTGGGAGCTGTGCA 737

QY 1577 GAAACTACTGCTGCTCCCTTTTTCACAGACATCTTTGGCCGAGAGCAGAGAAATGAAAAGC 1636

Db 738 GAAACTACTGCTGCTCCCTTTTTCACAGACATCTTTGGCCGAGAGCAGAGAAATGAAAAGC 797

QY 1637 CAGG 1640

Db 798 CAGG 801

RESULT 4
BX339313 1149 bp mRNA linear EST 02-MAY-2003
LOCUS BX339313 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CSOD1070YJ16 5-PRIME, mRNA sequence.
ACCESSION BX339313 GI:30335787
VERSION BX339313
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
1 (bases 1 to 1149)
Full-length cDNA libraries and normalization
Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6809.f For more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSOD1070DE08QPI&cluster=6809.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Paraday Avenue Genoscope sequence ID : CSOD1070DE08QPI.

FEATURES
Source Location/Qualifiers
1..1149
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSOD1070YJ16"
/cissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 240 a 285 c 298 g 41 others

ORIGIN
Query Match 34.0%; Score 773; DB 13; Length 1149;
Best Local Similarity 100.0%; Pred. No. 1,8e-123;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 GAGGGGAGCAGAGGGGAGTGTGGCGCTGGCGGGGTTTCGACGTGGGTTTCAGAG 152

Db 123 GAGGGGAGCAGAGGGGAGTGTGGCGCTGGCGGGGTTTCGACGTGGGTTTCAGAG 182

QY 153 CTGAGCCTTAAAGTGTGGAGAGTCCCTGGGCGCTCAGACAGGCGGCGGTGATC 212

Db 183 CTGAGCCTTAAAGTGTGGAGAGTCCCTGGGCGCTCAGACAGGCGGCGGTGATC 242

QY 213 ATCTCATCTTCTGGGTTACCTTGTGTTTATCGCATCTTTCTCAAG 272

Db 243 ATCTCATCTTCTGGGTTACCTTGTGTTTATCGCATCTTTCTCAAG 342

QY 273 GAGACTTACTATCCACTCTTCCATACCTTTACAGGCTCTCAATTGCTATTATTAAC 332

Db 303 GAGACTTACTATCCACTCTTCCATACCTTTACAGGCTCTCAATTGCTATTATTAAC 362

QY 333 TTGGAAACAGAGTACAGCTCCCTGCTGTGATGATGCTCAGTCTCCTTCTGCA 392

Db 363 TTGGAAACAGAGTACAGCTCCCTGCTGTGATGATGCTCAGTCTCCTTCTGCA 422

QY 393 CTAAATGGGCGGACATCACTGCGCTCTCACTACCTTTTGTCCAGATGGCTACCT 452

Db 423 CTAAATGGGCGGACATCACTGCGCTCTCACTACCTTTTGTCCAGATGGCTACCT 482

QY 453 CTGGCTGATTAATTAATCACTGCGGCACTAGATATCAATGAGCAATGCCACAT 512

Db 483 CTGGCTGATTAATTAATCACTGCGGCACTAGATATCAATGAGCAATGCCACAT 542

QY 513 TGTGTTCTGACTTTGAAGCTGATGTTGGTGGCTTTGACTCTTGAAGGAGGAAAGAT 572

Db 543 TGTGTTCTGACTTTGAAGCTGATGTTGGTGGCTTTGACTCTTGAAGGAGGAAAGAT 602

QY 573 CAGAAATTCCTGTCTCTGAGCAAGAAATATGCAATACGTGTGTTCTTCCCTGCTG 632

Db 603 CAGAAATTCCTGTCTCTGAGCAAGAAATATGCAATACGTGTGTTCTTCCCTGCTG 662

QY 633 GAAGTTGCTGTTTCTCTCTACTTCTATGAGGCTTTGTAGGCGCCAGTTCTCAATG 692

Db 663 GAAGTTGCTGTTTCTCTCTACTTCTATGAGGCTTTGTAGGCGCCAGTTCTCAATG 722

QY 693 AATCATCATGATGAAGCTGTGAGGAGAGTGAATGATCATACCAAGAAATACCAAC 752

Db 723 AATCATCATGATGAAGCTGTGAGGAGAGTGAATGATCATACCAAGAAATACCAAC 782

QY 753 AGCATATTCCTGCTCTCAAGCGCTGAGTGTGGCTTTTCTACTAGTGGCTACACA 812

Db 783 AGCATATTCCTGCTCTCAAGCGCTGAGTGTGGCTTTTCTACTAGTGGCTACACA 842

QY 813 CTGCTAGGCCCCCAGATCAGAGAGCTATCTCTCAGAGAGCTATGACAA 865

Db 843 CTGCTAGGCCCCCAGATCAGAGAGCTATCTCTCAGAGAGCTATGACAA 895

RESULT 5
B0917856 974 bp mRNA linear EST 20-AUG-2002
LOCUS B0917856
DEFINITION AGENCOUNT 8801447 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6311255
5', mRNA sequence.
ACCESSION B0917856
VERSION B0917856.1 GI:22332554
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
1 (bases 1 to 974)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
CDNA Library Preparation: Reggen, Invitrogen Corp
CDNA Library Arriayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL13734 row: n column: 24
High quality sequence stop: 755.
Location/Qualifiers
1..974
/organism="Mus musculus"
/mol_type="mRNA"

```

/db_xref="taxon:10090"
/clone="IMAGE:6311255"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_129"
/Note="Organ: olfactory epithelium; Vector:
pmv-SpOrf6.1.ccd; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dt. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH_MGC library."
BASE COUNT      199 a      260 c      248 g      267 t
ORIGIN

```

```

Query Match      33.7%; Score 767; DB 13; Length 974;
Best Local Similarity 100.0%; Pred. No. 2.1e-122;
Matches 767; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 68 CGAGTTAAGATGCGCTCTCTACAGCGAGGGGACGAGGAGCTGTGTGGCCCTGGCGG 127
DB 36 CGAGTTAAGATGCGCTCTCTACAGCGAGGGGACGAGGAGCTGTGTGGCCCTGGCGG 95
QY 128 GGTTCGAGTGGGGTTTCCAGAGCTAGCCCTTACAGTTGGGGAGCGTCCCTGGCGG 187
DB 96 GGTTCGAGTGGGGTTTCCAGAGCTAGCCCTTACAGTTGGGGAGCGTCCCTGGCGG 155
QY 188 GTCAAGACAGCGCGCTGCGCTGATCATCTCACTCTTCTGGTTACCCCTTGTCTTGT 247
DB 156 GTCAAGACAGCGCGCTGCGCTGATCATCTCACTCTTCTGGTTACCCCTTGTCTTGT 215
QY 248 TTATGGGCACTTACCTTTTCAAGAGACCTTACCTCATCCCTTCTCATACCTTTAC 307
DB 216 TTATGGGCACTTACCTTTTCAAGAGACCTTACCTCATCCCTTCTCATACCTTTAC 275
QY 308 AGGCGCTCAATGCTTATTTAATCTTGGAAACGAGTCTACCACTCCCTGCTGTAT 367
DB 276 AGGCGCTCAATGCTTATTTAATCTTGGAAACGAGTCTACCACTCCCTGCTGTAT 335
QY 368 TGTGCTTCAAGTCTCTCATCTTCACTAATGGGCGCACATCACTGCGCTCTCACTAC 427
DB 336 TGTGCTTCAAGTCTCTCATCTTCACTAATGGGCGCACATCACTGCGCTCTCACTAC 395
QY 428 CTTTTCCTTCCAGATGCGCTTACCTTGTGCTGGATATCTATTAACCTGCGCAACTA 487
DB 396 CTTTTCCTTCCAGATGCGCTTACCTTGTGCTGGATATCTATTAACCTGCGCAACTA 455
QY 488 CGATATCAAGTGGAGCAATGCAATGTTGTTGCTGATGAGTGGTGGTGGCTT 547
DB 456 CGATATCAAGTGGAGCAATGCAATGTTGTTGCTGATGAGTGGTGGTGGCTT 515
QY 548 TGAATCACTTGAAGGAGGAAAGATCAGAAATTCCTTCTGAGCAACAGAAATATGC 607
DB 516 TGAATCACTTGAAGGAGGAAAGATCAGAAATTCCTTCTGAGCAACAGAAATATGC 575
QY 608 CATACGTGTGTCTCTTCCCTGCTGGAAGTGTGCTTCTCTACTTATGGGCTT 667
DB 576 CATACGTGTGTCTCTTCCCTGCTGGAAGTGTGCTTCTCTACTTATGGGCTT 635
QY 668 CTTGAGGAGGCGCCAGTCTTCAATGATCACTAATGAGCTGGGAGAGAGTGTAT 727
DB 636 CTTGAGGAGGCGCCAGTCTTCAATGATCACTAATGAGCTGGGAGAGAGTGTAT 695
QY 728 TGAATCACTTGAAGGAGGAAAGATCAGAAAGATTCCTTCTGAGCAACAGAAATATGC 787
DB 696 TGAATCACTTGAAGGAGGAAAGATCAGAAAGATTCCTTCTGAGCAACAGAAATATGC 755
QY 788 CTTTTCCTTCAAGTGTGGCTTACACACTGCTGAGCGCCCAATCAAG 834
DB 756 CTTTTCCTTCAAGTGTGGCTTACACACTGCTGAGCGCCCAATCAAG 802

```

```

RESULT 5
BUS41837      BUS41837      916 bp      mRNA      linear      EST 13-SEP-2002
LOCUS      AGENCOURT_10325398 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6573588
DEFINITION

```

```

5', mRNA sequence.
BUS41837
BUS41837.1 GI:22852320
EST.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NIH-MGC http://mhc.nci.nih.gov/
1 (bases 1 to 916)
REFERENCE
AUTHORS
TITLES
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaabre@mail.nih.gov
Tissue Procurement: DCTD/DRP

```

```

cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LUCM2767 row: 1 column: 12
High quality sequence stop: 728.
Location/Qualifiers

```

FEATURES

```

source
1..916
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6573588"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/Note="Organ: prostate; Vector: pORF7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

```

BASE COUNT

```

189 a      249 c      221 g      257 t
ORIGIN

```

```

Query Match      33.0%; Score 751; DB 13; Length 916;
Best Local Similarity 100.0%; Pred. No. 1.2e-119;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 86 CTCACGAGGAGGAGCGAGGAGCTGTGTGCGCGGAGTTCGACATCGGCTT 145
DB 17 CTCACGAGGAGGAGCGAGGAGCTGTGTGCGCGGAGTTCGACATCGGCTT 76
QY 146 CAGAGAGCTGAGCCTTAAACAAGTTGGCGACGTCCTGGGCGCTCAGAAACAGGCGTGG 205
DB 77 CAGAGAGCTGAGCCTTAAACAAGTTGGCGACGTCCTGGGCGCTCAGAAACAGGCGTGG 136
QY 206 GCTGATCAATTCATCTTCTGAGGTTACCCCTTGTGTTTATGCGCATTAACCTTT 265
DB 137 GCTGATCAATTCATCTTCTGAGGTTACCCCTTGTGTTTATGCGCATTAACCTTT 196
QY 266 CTACAAGAGACCTACCTCATATCAGCTCTTCCATACCTTTACAGGCTCTCAATGCTTA 325
DB 197 CTACAAGAGACCTACCTCATATCAGCTCTTCCATACCTTTACAGGCTCTCAATGCTTA 256
QY 326 TTTTAACTTTGGAACAAGCTTACACACTGCTGCTGTGATGTGCTTCACTTCTCAT 385
DB 257 TTTTAACTTTGGAACAAGCTTACACACTGCTGCTGTGATGTGCTTCACTTCTCAT 316
QY 386 CTTGAGCTAATGGGCGGACATCATGCGGCTGCTCACTTATGCTTGGCTTCAAGATGGC 445
DB 317 CTTGAGCTAATGGGCGGACATCATGCGGCTGCTCACTTATGCTTGGCTTCAAGATGGC 376
QY 446 CTACCTTGTGCTGATATCTATTAACCTGCGACCGGCAACTAGATATCAATGACAT 505

```

Db 377 CTACCTTCTGCTGATGATCTATTACACTGCCACCGCACTACGATATCAAGTGACAT 436
QY 506 GCCACATTTGTTCTGACTTTGAAGCTGATTTGGCTGTGACTACTTTGACGAG 565
Db 437 GCCACATTTGTTCTGACTTTGAAGCTGATTTGGCTGTGACTACTTTGACGAG 496
QY 566 GAAAGATCAGAAATTCCTTCTGAGCAAGAAATATGCAATAGCTGTCTTCTTC 625
Db 497 GAAAGATCAGAAATTCCTTCTGAGCAAGAAATATGCAATAGCTGTCTTCTTC 556
QY 626 CCGCTGAGAGTGTGCTGCTTCTCTACTCTTAATGAGGCTTCTTGGTAGGCGCCAGT 685
Db 557 CCGCTGAGAGTGTGCTGCTTCTCTACTCTTAATGAGGCTTCTTGGTAGGCGCCAGT 616
QY 686 CTCATGAAATCACTACATGAGCTGTGAGGAGAGCTGATTTGACATACAGAAAGAT 745
Db 617 CTCATGAAATCACTACATGAGCTGTGAGGAGAGCTGATTTGACATACAGAAAGAT 676
QY 746 ACCAAACAGATATCTCTGCTCTCAAGCGCTGAGCTGTGCTTCTTACTAGTGG 805
Db 677 ACCAAACAGATATCTCTGCTCTCAAGCGCTGAGCTGTGCTTCTTACTAGTGG 736
QY 806 CTACACACTGCTGAGCCCGACATCACAGAA 836
Db 737 CTACACACTGCTGAGCCCGACATCACAGAA 767

RESULT 7
BG823157 800 bp mRNA linear EST 22-MAY-2001
LOCUS 602726506F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4866169 5',
DEFINITION mRNA sequence.
ACCESSION BG823157 GI:14170744
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgs.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgsbds-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LNCM1732 row: k column: 02
High quality sequence stop: 776.
Location/Qualifiers
1..800
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4866169"
/tissue_type="endocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_15"
/note="Organ: colon; Vector: pOT7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 161 a 218 c 199 g 222 t

Query Match 33.0%; Score 750; DB 12; Length 800;
Best Local Similarity 100.0%; Pred. No. 2e-119;
Matches 750; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 GGGGGGTCCTGTCGCGGAGTTAAGTGGCGTCTCAGCGGAGGAGGAGG 106
Db 4 GGGGGGTCCTGTCGCGGAGTTAAGTGGCGTCTCAGCGGAGGAGGAGG 63
QY 107 GACTGTGTGGGCGTCGCGGAGTTTTCAGAGTGGGTTTTCAGAGAGCTGAGCTTAACA 166
Db 64 GACTGTGTGGGCGTCGCGGAGTTTTCAGAGTGGGTTTTCAGAGAGCTGAGCTTAACA 123
QY 167 GTTGGGAGCGTCCCTGGGCGCTCAGAAAGCGCTGCGCTGATATCTCCATCTTCT 226
Db 124 GTTGGGAGCGTCCCTGGGCGCTCAGAAAGCGCTGCGCTGATATCTCCATCTTCT 183
QY 227 GGGTTACCCCTTGTGTTTATGCGATTACCTTTCTAAGAGAGACCTACCTCAT 286
Db 184 GGGTTACCCCTTGTGTTTATGCGATTACCTTTCTAAGAGAGACCTACCTCAT 243
QY 287 CCACCTCTTCATACCTTTTACAGGCGCTCTCAATTGCTTATTTTGAACACAGCT 346
Db 244 CCACCTCTTCATACCTTTTACAGGCGCTCTCAATTGCTTATTTTGAACACAGCT 303
QY 347 CTACCACTCCCTGCTGTGATTTGCTTCACTTCTTCCATTCGACTAATGGCGGCAC 406
Db 304 CTACCACTCCCTGCTGTGATTTGCTTCACTTCTTCCATTCGACTAATGGCGGCAC 363
QY 407 CATCACTGCGGCTCTCACTACCTTTGCTTCCAGATGGGCTTCTGCTGCTGATACCTA 466
Db 364 CATCACTGCGGCTCTCACTACCTTTGCTTCCAGATGGGCTTCTGCTGCTGATACCTA 423
QY 467 TTACACTGCGACCGGCACTACGATATCAAGTGAATGCCAATGTGTTCTGACTT 526
Db 424 TTACACTGCGACCGGCACTACGATATCAAGTGAATGCCAATGTGTTCTGACTT 483
QY 527 GAAGCTGATTTGTTGGCTGTGACTTATGACGAGGAGAAAGATCAGAAATTCCTTCT 586
Db 484 GAAGCTGATTTGTTGGCTGTGACTTATGACGAGGAGAAAGATCAGAAATTCCTTCT 543
QY 587 CTCGTAGCAACAGAAATATGCGCATACGATGATCAATGCTGCTGCTGCTGCTGCTG 646
Db 544 CTCGTAGCAACAGAAATATGCGCATACGATGATCAATGCTGCTGCTGCTGCTGCTG 603
QY 647 CTCCTACTTCTATGGGCTCTTCTGTAGGCGCCAGCTTCTCAATGATCACTACATGAA 706
Db 604 CTCCTACTTCTATGGGCTCTTCTGTAGGCGCCAGCTTCTCAATGATCACTACATGAA 663
QY 707 GCTGTGCGAGGAGAGCTGATTTGACATACCAAGAAATATCCAAACAGATCTTCTGC 766
Db 664 GCTGTGCGAGGAGAGCTGATTTGACATACCAAGAAATATCCAAACAGATCTTCTGC 723
QY 767 TCTCAAGCGCTGAGTGTGGGCTTTTCTA 796
Db 724 TCTCAAGCGCTGAGTGTGGGCTTTTCTA 753

RESULT 8
CA489207 911 bp mRNA linear EST 14-NOV-2002
LOCUS CA489207
DEFINITION AGENCOURT_10809550 MAPCL Homo sapiens cDNA clone IMAGE:6721545 5',
mRNA sequence.
ACCESSION CA489207 GI:24951998
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgs.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE

JOURNAL COMMENT

Unpublished
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: Kristi A. Eglund, Ira Pastan
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLNL4282 row: n column: 09
High quality sequence stop: 657.
Location/Qualifiers

FEATURES source

1. 911
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6721545"
/cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231, HTERT-HME1",
"LNCaP"
/lab_host="EMD10B"
/clone_1lb="MAPL"
/note="Vector: pCMV-Sport6; Site 1: EcoRV, Site 2: Not I; Subtracted with brain, liver, lung, kidney and muscle. Directionally cloned. Priming method: oligo-dt. Average insert size: 1800 bp. Library amplification: 26,000 fold. Kristi A. Eglund, James J. Vincent, Robert Strausberg, Bungkook Lee & Ira Pastan: Discovery of new breast cancer genes encoding membrane and secreted proteins. Manuscript submitted."

BASE COUNT 219 a 253 c 203 g 234 t 2 others

ORIGIN

Query Match 32.9%; Score 747; DB 14; Length 911;
Best Local Similarity 100.0%; Pred. No. 5.7e-119;
Matches 747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 764 TCGCTCAAGCGCTGAGTCTGGGCTTTCTACTAGTGGGCTACACACTGCTGAGCCC 823
Db 1 TCGCTCAAGCGCTGAGTCTGGGCTTTCTACTAGTGGGCTACACACTGCTGAGCCC 60
QY 824 CCACATCACAAGAACTATCTCTCTCACTGAAGATATGACAAACACCCCTTCTGTTCCG 883
Db 61 CCACATCACAAGAACTATCTCTCTCACTGAAGATATGACAAACACCCCTTCTGTTCCG 120
QY 884 CTGCATGATCATGCTGATCTGGGGCAAGTTGTGCTGTACAAATATGTCACCTGTTGGCT 943
Db 121 CTGCATGATCATGCTGATCTGGGGCAAGTTGTGCTGTACAAATATGTCACCTGTTGGCT 180
QY 944 GGTCAACAAGAGATATGATTTTGAAGGCTGGGCTTCAATGCTTTGAAGAAAGG 1003
Db 181 GGTCAACAAGAGATATGATTTTGAAGGCTGGGCTTCAATGCTTTGAAGAAAGG 240
QY 1004 CAAGCAAAAGTGGATGCTGTGTCACAAATGAAGGTGTGCTTTGAAGAAAGG 1063
Db 241 CAAGCAAAAGTGGATGCTGTGTCACAAATGAAGGTGTGCTTTGAAGAAAGG 300
QY 1064 CTTCCTGACACATGCTCATTTTCAATCAACACCAAGCTGGGTGGCCGCTAAT 1123
Db 301 CTTCCTGACACATGCTCATTTTCAATCAACACCAAGCTGGGTGGCCGCTAAT 360
QY 1124 CTTCAAACGACTCAATGCTCTTGAAGAAATCTCTGACGGGTCTCTGTTGCTATT 1183
Db 361 CTTCAAACGACTCAATGCTCTTGAAGAAATCTCTGACGGGTCTCTGTTGCTATT 420
QY 1184 CCTGGCCCTCTGTCAGCGCTGTCATCAGAGATACCTGTGCTGCTTCAAGTGAATTCCT 1243
Db 421 CCTGGCCCTCTGTCAGCGCTGTCATCAGAGATACCTGTGCTGCTTCAAGTGAATTCCT 480
QY 1244 CATGTGATATGGAAGAAGCAGGCTGCCAGGCTCATTCAAGAGAGCCCACTGAGCA 1303
Db 481 CATGTGATATGGAAGAAGCAGGCTGCCAGGCTCATTCAAGAGAGCCCACTGAGCA 540

QY 1304 GCTGCCGCGCATTAATCTCTCTCCAGCCCTTCTACTATTGTGTGAACAGACCATTCACATG 1363
Db 541 GCTGCCGCGCATTAATCTCTCTCCAGCCCTTCTACTATTGTGTGAACAGACCATTCACATG 600
QY 1364 GCTCTTCATGGGTACTCTCCATGACTGCTTCTGCTCTTCACTGAGGACAAATGGCTTAA 1423
Db 601 GCTCTTCATGGGTACTCTCCATGACTGCTTCTGCTCTTCACTGAGGACAAATGGCTTAA 660
QY 1424 GGTGTATTAATCACTATTTCTGTCGACATCTTCTTCTGAGCTACTATTCATATT 1483
Db 661 GGTGTATTAATCACTATTTCTGTCGACATCTTCTTCTGAGCTACTATTCATATT 720
QY 1484 GCCTTATATTCACAAAGCAATGTGCC 1510
Db 721 GCCTTATATTCACAAAGCAATGTGCC 747

RESULT 9
BM557200 1053 bp mRNA linear EST 20-FEB-2002
LOCUS BM557200
DEFINITION AGENCOURT 6579051 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5466619
5' mRNA sequence.
ACCESSION BM557200
VERSION BM557200.1 GI:18799007
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
1 (bases 1 to 1053)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: DCTD/DRP

CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLCM1967 row: e column: 20
High quality sequence stop: 677.
Location/Qualifiers

FEATURES source

1. 1053
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5466619"
/tissue_type="melanotic, melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI, Site 2: EcoRI; CDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC library."

BASE COUNT 230 a 313 c 237 g 267 t 6 others

ORIGIN

Query Match 32.6%; Score 741; DB 12; Length 1053;
Best Local Similarity 99.8%; Pred. No. 5.4e-118;
Matches 841; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1092 ATCAACACCAAGCGCTGGGCTGCGCTACATCTTCAACGACTGAATCTTGAAT 1151
Db 1 ATCAACACCAAGCGCTGGGCTGCGCTACATCTTCAACGACTGAATCTTGAAT 60
QY 1152 AAGAAGCTCTCAGAGGTCTCTGTTGCTATTCCTGCGCTGACAGGCTGACATCA 1211

Db 61 AAGAACTCTCTAGAGGCTCTCTGTCATTTCTGAGCCCTCTGAGAGGCTGCACTTCA 120
1212 GGATACCTGTGTCTTCCAGATGAAATTCCTCATTTGTTATTTGGAAGACAGCTGCC 1271
121 GGATACCTGTGTCTTCCAGATGAAATTCCTCATTTGTTATTTGGAAGACAGGCTGCC 180
1272 AGGCTATTCAAGAGAGCCCACTGAGCAAGCTGAGCCCACTTACTGTCTTCAGGCC 1331
181 AGGCTATTCAAGAGAGCCCACTGAGCAAGCTGAGCCCACTTACTGTCTTCAGGCC 240
1332 TTCTACTATTGTGTCAGACAGCACTGAGCTCTTCAATGAGTTACTCATGAGCTGCC 1391
241 TTCTACTATTGTGTCAGACAGCACTGAGCTCTTCAATGAGTTACTCATGAGCTGCC 300
1392 TTCTGCTCTTCAAGTGGACAAATGAGCTTAAAGTATTAATTCATTTCTTCTTGGC 1451
301 TTCTGCTCTTCAAGTGGACAAATGAGCTTAAAGTATTAATTCATTTCTTCTTGGC 360
1452 CACATCTTCTTCTGAGCCCTACTATTCATATTTGCTTATTTCAAGAAAGATGTTGCCA 1511
361 CACATCTTCTTCTGAGCCCTACTATTCATATTTGCTTATTTCAAGAAAGATGTTGCCA 420
1512 AGGAAAGAGAGTTAAAGAGATGAAATTAATTCATTTCTGAGCTGAGCTGCCAGCTG 1571
421 AGGAAAGAGAGTTAAAGAGATGAAATTAATTCATTTCTGAGCTGAGCTGCCAGCTG 480
1572 GTGCAAAACTACTGCTCTCTCTTTTCAAGCACTCTCTTTGCCAGAGAGAGATGGA 1631
481 GTGCAAAACTACTGCTCTCTCTTTTCAAGCACTCTCTTTGCCAGAGAGAGATGGA 540
1632 AAGGCAAGGAGATGGAAGATGATGCTTCCAGCTGAGCTCTGCTGCTGAGCCAGAGCTT 1691
541 AAGGCAAGGAGATGGAAGATGATGCTTCCAGCTGAGCTCTGCTGCTGAGCCAGAGCTT 600
1692 CATTGGGGGCAAGAGGGAATTTTCTTTGGAAGAGGCTCTTCTGCTTGTCAAGCCAGC 1751
601 CATTGGGGGCAAGAGGGAATTTTCTTTGGAAGAGGCTCTTCTGCTTGTCAAGCCAGC 660
1752 TGGAAATGACAGTGGCGGAGTCTCAGCTCAGCCGCAACCTCTGCTGAGGTTCAAGTAT 1811
661 TGGAAATGACAGTGGCGGAGTCTCAGCTCAGCCGCAACCTCTGCTGAGGTTCAAGTAT 720
1812 TTCTGCTCTGAGCTCTCCAGTATGCTGGGAATACAGGCAAGCCAGCTGAGCTTAT 1871
721 TTCTGCTCTGAGCTCTCCAGTATGCTGGGAATACAGGCAAGCCAGCTGAGCTTAT 780
1872 TTGTGATTTTCACTAGAAAGGATTTCAACAGCTGAGCTGAGCTGAGCTGAGCTG 1931
781 TTGTGATTTTCACTAGAAAGGATTTCAACAGCTGAGCTGAGCTGAGCTGAGCTG 840
1932 ACC 1934
841 ACC 843

RESULT 10.
AL557774 1201 bp mRNA linear EST 31-MAY-2003
LOCUS AL557774 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0D003YA18 5-PRIME, mRNA sequence.
ACCESSION AL557774
VERSION AL557774.2 GI:31279574
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 15, 2001 this sequence version replaced gi:12901710.

Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seq@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6809.f. For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D003BA09QPI&cluster=6809.f. Contact :
Feng Liang Email : fliang@life.techn.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0D003BA09QPI.
FEATURES
Source
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D003YA18"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
10-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 229 a 315 c 282 g 304 t 71 others
ORIGIN
Query Match 32.2%; Score 733; DB 9; Length 1201;
Best Local Similarity 100.0%; Pred. No. 1,1e-116;
Matches 733; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
48 GGGGGTCCCTGTGGGCTCCGAGTTAAGATGCGCTCTCAGCGAGGGGAGCGAGGG 107
32 GGGGGTCCCTGTGGGCTCCGAGTTAAGATGCGCTCTCAGCGAGGGGAGCGAGGG 91
108 ACTGTGTGGGCTGTGGGCTGTGCACTGCGGTTTCCAGAGCTGAGCTTAAACAG 167
92 ACTGTGTGGGCTGTGGGCTGTGCACTGCGGTTTCCAGAGCTGAGCTTAAACAG 151
168 TTGGCAGCTCCCTGTGGGCTGTGCAAGAGGCGCTGCGTATCATCTCATCTTCG 227
152 TTGGCAGCTCCCTGTGGGCTGTGCAAGAGGCGCTGCGTATCATCTCATCTTCG 211
228 GGTACCCCTTGTGCTTTATGCGCATTAACCTTTTCAAGAGACTTACCTCATC 287
212 GGTACCCCTTGTGCTTTATGCGCATTAACCTTTTCAAGAGACTTACCTCATC 271
288 CACCTCTTCATACCTTTACAGGCTCTCAATGCTTATTTTAACTTTGGAACAGCTC 347
272 CACCTCTTCATACCTTTACAGGCTCTCAATGCTTATTTTAACTTTGGAACAGCTC 331
348 TACCACTCCCTGTGTATGATGCTTCAAGTTCCTCATCTTCGATTAATGAGCGCAGC 407
332 TACCACTCCCTGTGTATGATGCTTCAAGTTCCTCATCTTCGATTAATGAGCGCAGC 391
408 ATCACTGCGCTCTCACTACCTTTTGTCCAGATGAGCTTACCTTGTGCTGATCTAT 467
392 ATCACTGCGCTCTCACTACCTTTTGTCCAGATGAGCTTACCTTGTGCTGATCTAT 451
468 TACACTGCCACCGGCACTACATCAAGTGGACAAATGCGCATTTGTGTGACTTTG 527
452 TACACTGCCACCGGCACTACATCAAGTGGACAAATGCGCATTTGTGTGACTTTG 511
528 AAGCTATTGTTGTGCTGTGCTACTCTTTCAGAGGAGGAAAGATCAGATTCTTGTGTC 587
512 AAGCTATTGTTGTGCTGTGCTACTCTTTCAGAGGAGGAAAGATCAGATTCTTGTGTC 571
588 TCTGAGCAACAGAAATATGACATGAGTGTGTTCTTCCCTGTGGAAGTTGCTGTTTC 647
572 TCTGAGCAACAGAAATATGACATGAGTGTGTTCTTCCCTGTGGAAGTTGCTGTTTC 631

QY 648 TCTACTTCTATGGGCGCTTCTGTAGAGGCCAGTTCATGATCATCATGAG 707
DB 632 TCTACTTCTATGGGCGCTTCTGTAGAGGCCAGTTCATGATCATCATGAG 691
QY 708 CTGGTCAGGAGAGCTGATTTGACATACCGAGAAAGATACCAACAGCATATTCTGCT 767
DB 692 CTGGTCAGGAGAGCTGATTTGACATACCGAGAAAGATACCAACAGCATATTCTGCT 751
QY 768 CTCAGCGCCTGA 780
DB 752 CTCAGCGCCTGA 764

RESULT 11
BG753831 887 bp mRNA linear EST 15-MAY-2001
LOCUS 602733053F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4876525 5',
DEFINITION mRNA sequence.
ACCESSION BG753831 GI:14064484
VERSION BG753831
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 887)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNL at:
http://image.llnl.gov
Plate: L1CM1759 row: j column: 14
High quality sequence stop: 814.
Location/Qualifiers
1. 887
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4876525"
/issue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 43"
/note="Organ: eye; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library. |"

BASE COUNT 208 a 250 c 203 g 226 t
ORIGIN

Query Match 31.5%; Score 715; DB 10; Length 887;
Best Local Similarity 99.9%; Pred. No. 1.7e-113;
Matches 765; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1020 GCGTGTGCGCAATGAAGGTGTGCTTTGAAACAAACCCCGCTTCACTGGACCAATT 1079
DB 1 GCGTGTGCGCAATGAAGGTGTGCTTTGAAACAAACCCCGCTTCACTGGACCAATT 60
QY 1080 GCGTCAATCAACATCAACCAACGCGTGGTGGCCGCTTCACTTTCAACAGACTCAAG 1139
DB 61 GCGTCAATCAACATCAACCAACGCGTGGTGGCCGCTTCACTTTCAACAGACTCAAG 120
QY 1140 TTCCTTGAATAAAGACTCTCTCAGGGTCTCTCGTTCATTCCTGAGCCCTTGGCAC 1199

DB 122 TTCCTTGAATAAAGACTCTCTCAGGGTCTCTCGTTCATTCCTGAGCCCTTGGCAC 180
QY 1200 GCGTGTGCGCAATGAAGGTGTGCTTTGAAACAAACCCCGCTTCACTGGACCAATT 1259
DB 181 GCGTGTGCGCAATGAAGGTGTGCTTTGAAACAAACCCCGCTTCACTGGACCAATT 240
QY 1260 AGACAGGCTGCCAGGCTCATTTGAAGAGAGCCCACTGAGCAAGCTGGCCGCTTACT 1319
DB 241 AGACAGGCTGCCAGGCTCATTTGAAGAGAGCCCACTGAGCAAGCTGGCCGCTTACT 300
QY 1320 GTTCCTCCAGCCCTTCTACTATTGTGTGCAACAGACCATTCAGTGCCTTTCATGGATTAC 1379
DB 301 GTTCCTCCAGCCCTTCTACTATTGTGTGCAACAGACCATTCAGTGCCTTTCATGGATTAC 360
QY 1380 TCCATGACTGCTTCTGCTCTTTCACGTGGGCAAAATGCTTAAGGTATTAATTCATC 1439
DB 361 TCCATGACTGCTTCTGCTCTTTCACGTGGGCAAAATGCTTAAGGTATTAATTCATC 420
QY 1440 TATTTCTGGGCAACATCTTCTCCGAGCCCTACTATTAATGCTTATTAATTCACAA 1499
DB 421 TATTTCTGGGCAACATCTTCTCCGAGCCCTACTATTAATGCTTATTAATTCACAA 480
QY 1500 GCAATGTGTCGAAGAAAGAAAGATTAAAGAAAGATTAATCCATTTCCCTGTGGCC 1559
DB 481 GCAATGTGTCGAAGAAAGAAAGATTAAAGAAAGATTAATCCATTTCCCTGTGGCC 540
QY 1560 TGTGCGGAGCTGTGCAAACTACTGCTCCCTTTTCAACAGACTCTTGTGCCCCAGA 1619
DB 541 TGTGCGGAGCTGTGCAAACTACTGCTCCCTTTTCAACAGACTCTTGTGCCCCAGA 600
QY 1620 GCAGAAATGGAAGAAAGCGGAGGTGAATGCAATGCTTCAGTGTGCGCTGTGCTC 1679
DB 601 GCAGAAATGGAAGAAAGCGGAGGTGAATGCAATGCTTCAGTGTGCGCTGTGCTC 660
QY 1680 CAGCCAGCTTCATTTGGGGCCAAAGGGAAATCTTTTGGAGAAAGCGCTTGTCTT 1739
DB 661 CAGCCAGCTTCATTTGGGGCCAAAGGGAAATCTTTTGGAGAAAGCGCTTGTCTT 720
QY 1740 TGTCAACCAAGCTGGAATGAGTGGCGGATCTCAGCTCACCGCAA 1785
DB 721 TGTCAACCAAGCTGGAATGAGTGGCGGATCTCAGCTCACCGCAA 766

RESULT 12
B1760522 847 bp mRNA linear EST 25-SEP-2001
LOCUS 603045066F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5185120 5',
DEFINITION mRNA sequence.
ACCESSION B1760522
VERSION B1760522.1 GI:15752100
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 847)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNL at:
http://image.llnl.gov
Plate: L1AM11462 row: d column: 17
High quality sequence stop: 807.
Location/Qualifiers
1. 847
source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5185120"
/lab_host="DH10B"
/clone_1ib="NIH MGC 116"
/note="Organ: pooled colon, kidney, stomach; Vector:
PCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 3 colons, age 26 yo male, 49 yo
female, 71 yo male colon; 46 yo male kidney, and pool of 2
stomachs, 62 yo male and 70 yo female. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.4 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
023. Note: this is a NIH MGC Library."
BASE COUNT      170 a      236 c      210 g      231 t
ORIGIN

```

```

Query Match      31.3%; Score 711; DB 12; Length 847;
Best Local Similarity 100.0%; Pred. No. 8.4e-113;
Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

47  GGGGGGCTCTGCGGCTCCGAGTAAAGTGGCTCTCAGCGGAGGGGACGAGG 106
    |||||||
17  GGGGGGCTCTGCGGCTCCGAGTAAAGTGGCTCTCAGCGGAGGGGACGAGG 76
    |||||||
107 GACTGATGAGCGCTGCGGCGGCTTCGAGTGGGTTTCAGAGCTGAGCTTAACA 166
    |||||||
77  GACTGATGAGCGCTGCGGCGGCTTCGAGTGGGTTTCAGAGCTGAGCTTAACA 136
    |||||||
167 GTTGGCGAGCTCCCTGGGCGCTGAGAACAGCGCTCGGCTGATCATCTTCT 226
    |||||||
137 GTTGGCGAGCTCCCTGGGCGCTGAGAACAGCGCTCGGCTGATCATCTTCT 196
    |||||||
227 GGGTTACCCCTTGTGTTTATGGCAATACCTTTTACAGAGAGCACTACCTCAT 286
    |||||||
197 GGGTTACCCCTTGTGTTTATGGCAATACCTTTTACAGAGAGCACTACCTCAT 256
    |||||||
287 CCACCTCTTCATACCTTTTACAGAGCGCTCAATGCTTATTTTAACTTGAAC 346
    |||||||
257 CCACCTCTTCATACCTTTTACAGAGCGCTCAATGCTTATTTTAACTTGAAC 316
    |||||||
347 CTACCACTCCCTGCTGTGATGTTGCTTCAAGTTCCTCATCTTGAATATGG 406
    |||||||
317 CTACCACTCCCTGCTGTGATGTTGCTTCAAGTTCCTCATCTTGAATATGG 376
    |||||||
407 CATCAGCGGCTCTCACTCTTGTGCTTCCAGATGGCTTACCTTGGCTGGATCA 466
    |||||||
377 CATCAGCGGCTCTCACTCTTGTGCTTCCAGATGGCTTACCTTGGCTGGATCA 436
    |||||||
467 TTACACTGCCAGCGCAACTACGATATCAAGTGGACATGCCAATGTTGCTGACT 526
    |||||||
437 TTACACTGCCAGCGCAACTACGATATCAAGTGGACATGCCAATGTTGCTGACT 496
    |||||||
527 GAAAGTGTGTTGTTGCTGTGACTTCTTGAAGGAGGAAAGATCAGAATTCCTT 586
    |||||||
497 GAAAGTGTGTTGTTGCTGTGACTTCTTGAAGGAGGAAAGATCAGAATTCCTT 556
    |||||||
587 CTCTGAGCAACAGAAATATGCAATGAGTGGTTCCTTCCCTGTGGAAGTGTGTT 646
    |||||||
557 CTCTGAGCAACAGAAATATGCAATGAGTGGTTCCTTCCCTGTGGAAGTGTGTT 616
    |||||||
647 CTCTTACTTATAGGAGCTTCTTGTAGAGGCCCACTTCTCAATGATCACTATGAA 706
    |||||||
617 CTCTTACTTATAGGAGCTTCTTGTAGAGGCCCACTTCTCAATGATCACTATGAA 676
    |||||||
707 GCTGTGTCAGAGAGGCTGATGACATACAGAAAGATACCAACAGCAT 757
    |||||||
677 GCTGTGTCAGAGAGGCTGATGACATACAGAAAGATACCAACAGCAT 727
    |||||||

```

```

RESULT 13
AL553069      1201 bp      mRNA      linear      EST 31-MAY-2003
LOCUS
DEFINITION
AL553069 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1072Y019 5-PRIME, mRNA sequence.
ACCESSION
AL553069
VERSION
AL553069.2 GI:31274883
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 15, 2001 this sequence version replaced gi:12892559.
COMMENT
Genoscope - Centre National de Sequenage
BP 191 91006 EVRY cedex - France
Email: segr@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6809.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1072AH10P1&cluster=6809.f. Contact :
Feng Liang Email : fliang@life.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0D1072AH10P1.
Location/Qualifiers
1..1201

```

FEATURES

```

source
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1072Y019"
/issue_type="PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the PCMVSPORT 6 vector. Library was normalized."

```

```

BASE COUNT      288 a      273 c      274 g      309 t
ORIGIN
Query Match      31.1%; Score 708; DB 9; Length 1201;
Best Local Similarity 100.0%; Pred. No. 2.1e-112;
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

512 TTGTGTTGACTTGAAGCTGATGTTGGTGTGACTACTTGAACGAGGAAAGA 571
    |||||||
63  TTGTGTTGACTTGAAGCTGATGTTGGTGTGACTACTTGAACGAGGAAAGA 122
    |||||||
572 TCAGAAATTCCTTGTCTCTGAGCAAGAAATATGCAATCGTGTTCCTTCC 631
    |||||||
123 TCAGAAATTCCTTGTCTCTGAGCAAGAAATATGCAATCGTGTTCCTTCC 182
    |||||||
632 GGAAGTGTGTTTCTCTACTTCTATGAGGCTTCTTGTGAGGCCCCCAAT 691
    |||||||
183 GGAAGTGTGTTTCTCTACTTCTATGAGGCTTCTTGTGAGGCCCCCAAT 242
    |||||||
692 GAATCATCATAGAGTGTGAGGAGAGCTGATGACATACAGAAAGATACCAA 751
    |||||||
243 GAATCATCATAGAGTGTGAGGAGAGCTGATGACATACAGAAAGATACCAA 302
    |||||||
752 CAGCATCATCTCTGCTCTCAAGCGCTGAGTGTGAGCTTTTCTACTAGTGGCTA 811
    |||||||
303 CAGCATCATCTCTGCTCTCAAGCGCTGAGTGTGAGCTTTTCTACTAGTGGCTA 362
    |||||||
812 ACTGCTCAGCCCCCATCATCAGAAAGCTATCTCTCACTGAAGACTATACCA 871
    |||||||
363 ACTGCTCAGCCCCCATCATCAGAAAGCTATCTCTCACTGAAGACTATACCA 422
    |||||||
872 CTCTGTGTTCCGCTGATGATGCTGATCTGGGCAAGTTGTGCTGTAATAATGT 931
    |||||||

```



```

Db      423 CTTCTGCTCCCGCTGATGATGCTGAGTCTGGGGCAAGTTTGCTTACAAATATGT 482
Qy      932 CACCTGTTGGCTGTGACAGAAAGATGATGATTTTGAAGGGCTGGGCTTCAATGGCTT 991
Db      483 CACCTGTTGGCTGTGACAGAAAGATGATGATTTTGAAGGGCTGGGCTTCAATGGCTT 542
Qy      992 TGAAGAAAAGGCAAGCAAGTGGATGCTGTGGCAATGATGAAGTGGCTTTTGA 1051
Db      543 TGAAGAAAAGGCAAGCAAGTGGATGCTGTGGCAATGATGAAGTGGCTTTTGA 602
Qy      1052 AACAAACCCCGCTTCACTGGGACCATGCTCTTCAATCAATCAACCAACGCTGGGT 1111
Db      603 AACAAACCCCGCTTCACTGGGACCATGCTCTTCAATCAATCAACCAACGCTGGGT 662
Qy      1112 GGGCCGCTACATCTTCAAGCACTCAAGTCTTCTGAAATTAAGAACTCTTCAGAGTCT 1171
Db      663 GGGCCGCTACATCTTCAAGCACTCAAGTCTTCTGAAATTAAGAACTCTTCAGAGTCT 722
Qy      1172 CTCGTTGCTATTCTGCGCCCTCTGGACGCGCTGCATCAGATACCT 1219
Db      723 CTCGTTGCTATTCTGCGCCCTCTGGACGCGCTGCATCAGATACCT 770

```

```

RESULT 14
LOCUS   BM553135 1097 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT_6572558 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5467062
5', mRNA sequence.
ACCESSION BM553135
VERSION   BM553135.1 GI:18791600
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     NIH-MGC http://mgs.nci.nih.gov/.
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1968 row: h column: 07
High quality sequence stop: 666.
Location/Qualifiers

```

```

FEATURES
source
1..1097

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5467062"
/issue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1b="NIH MGC 41"
/notes="Organ: skin; Vector: pOT7; Site 1: XhoI; Site 2:
ECORI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC library."

```

```

BASE COUNT 244 a 315 c 254 g 275 t 9 others
Query Match 31.1%; Score 707; DB 12; Length 1097;
Best Local Similarity 100.0%; Pred. No. 3.3e-112;

```

```

Matches 707; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1092 ATCAACACCAACGCGCTGGGTGGCCCGCTACATCTTCAAAAGCACTGAAGTCTTGGAAAT 1151
Db      1 ATCAACACCAACGCGCTGGGTGGCCCGCTACATCTTCAAAAGCACTGAAGTCTTGGAAAT 60
Qy      1152 AAGAAGCTCTCAGAGGTCTCTCGTTGCTATTCCTGGCCCTTGGCAGGCTGGACATCA 1211
Db      61 AAGAAGCTCTCAGAGGTCTCTCGTTGCTATTCCTGGCCCTTGGCAGGCTGGACATCA 120
Qy      1212 GATACCTGCTGCTCTTCCAGATGGAATTCCTCATTTGTTATTTGGAAGACAGCTGGC 1271
Db      121 GATACCTGCTGCTCTTCCAGATGGAATTCCTCATTTGTTATTTGGAAGACAGCTGGC 180
Qy      1272 AGGCTCATTCAGAGAGGCCCACTTGAGCAAGCTGGCGGCATTAAGTCTCTCAGAGCC 1331
Db      181 AGGCTCATTCAGAGAGGCCCACTTGAGCAAGCTGGCGGCATTAAGTCTCTCAGAGCC 240
Qy      1332 TTCTACTATTTGGTGCACAGACATCCACTGGCTCTTCATGAGTTACTCCATGACTGCC 1391
Db      241 TTCTACTATTTGGTGCACAGACATCCACTGGCTCTTCATGAGTTACTCCATGACTGCC 300
Qy      1392 TTCTGCTCTTTCACCTGGGACCAATGGCTTAAAGTGTATTAATTCATCTATTCTTGGC 1451
Db      301 TTCTGCTCTTTCACCTGGGACCAATGGCTTAAAGTGTATTAATTCATCTATTCTTGGC 360
Qy      1452 CACATCTTCTTCCCTAGGCTTACTATTCATTTGGCTTATATTCACAAAGCATGGTGCCA 1511
Db      361 CACATCTTCTTCCCTAGGCTTACTATTCATTTGGCTTATATTCACAAAGCATGGTGCCA 420
Qy      1512 AGGAAGAGAAAGTTAAAGAGATGGAATATCATTTCCCTGGTGGCTGTGGGACTG 1571
Db      421 AGGAAGAGAAAGTTAAAGAGATGGAATATCATTTCCCTGGTGGCTGTGGGACTG 480
Qy      1572 GTGCAGAACTACTGCTCTCTCTTTTTCACAGCACTCTTTGCCCCAGACAGATGGA 1631
Db      481 GTGCAGAACTACTGCTCTCTCTTTTTCACAGCACTCTTTGCCCCAGACAGATGGA 540
Qy      1632 AAGCCAGGAGAGTGGAAATGATGCTTCCAGCTGTGCTTCTGCTCCAGCAAGTCTT 1691
Db      541 AAGCCAGGAGAGTGGAAATGATGCTTCCAGCTGTGCTTCTGCTCCAGCAAGTCTT 600
Qy      1692 CATTGGGGGCCAAAGGGGAACTTTTGTGAGAAGGGCTCTGCTTGTGACCAACGC 1751
Db      601 CATTGGGGGCCAAAGGGGAACTTTTGTGAGAAGGGCTCTGCTTGTGACCAACGC 660
Qy      1752 TGAATGCAAGTGGCGGATCTCAGCTCACCGCACTCCACCTCTCG 1798
Db      661 TGAATGCAAGTGGCGGATCTCAGCTCACCGCACTCCACCTCTCG 707

```

```

RESULT 15
LOCUS   BU597281 827 bp mRNA linear EST 20-SEP-2002
DEFINITION AGENCOURT_8966025 NIH_MGC_142 Homo sapiens cDNA clone IMAGE:4454096
5', mRNA sequence.
ACCESSION BU597281
VERSION   BU597281.1 GI:23249040
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens

```

```

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     NIH-MGC http://mgs.nci.nih.gov/.
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: NCI
CDNA Library Preparation: Michael Brownstein Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

```

```

BASE COUNT 244 a 315 c 254 g 275 t 9 others
Query Match 31.1%; Score 707; DB 12; Length 1097;
Best Local Similarity 100.0%; Pred. No. 3.3e-112;

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>

Plate: LNCM2630 row: j column: 17
High quality sequence stop: 559.
Location/Qualifiers

FEATURES

source

```

1.827
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6454096"
/tissue_type="mixed (pool of 40 RNAs)"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH_MGC_142"
/notes="Vector: pDNR-LIB; Site 1: SfiI (ggccatcattggcc); Site 2: SfiI (ggcgctcgccg); Double-stranded cDNA was prepared from a pool of 40 cell line polyA+ RNAs (bladder - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon - 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%, kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were used in cloning as follows:
5'-AAGCAGTGTATCAAGCAGAGTGGCCATTAGGCGCGG-3' and
5'-ATTCTAGAGGCCGAGCGCGCCGACATG-dt(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the >0.5 kb size fraction (other fractions present in NIH_MGC_141). Library created in the laboratory of M. Brownstein (NIH, NIH). Note: this is a NIH MGC library."
BASE COUNT      185 a      238 c      200 g      201 t      3 others
ORIGIN

```

Query Match 30.9%; Score 703; DB 13; Length 827;
Best Local Similarity 100.0%; Pred. No. 2e-11;
Matches 703; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1248 GTTATTGTGGAAGACAGGCTGCCAGGCTTATTCAGAGAGCCCGCCCTGAGCAAGCTG 1307
DB 3 GTTATTGTGGAAGACAGGCTGCCAGGCTTATTCAGAGAGCCCGCCCTGAGCAAGCTG 62
QY 1308 GCGGCATTACTGCTCCCGACGCCCTTACTATTGTGTGACACAGACCATCACTGGCTC 1367
DB 63 GCGGCATTACTGCTCCCGACGCCCTTACTATTGTGTGACACAGACCATCACTGGCTC 122
QY 1368 TTGATGGGTACTCATGACTGCTTCTGCTCTTCACTGGGACAATGGCTTAAGTG 1427
DB 123 TTGATGGGTACTCATGACTGCTTCTGCTCTTCACTGGGACAATGGCTTAAGTG 182
QY 1428 TATTAATCCATCTATTCTTGGCCACATCTTCTCTGAGCTACTATTGATTTGGCT 1487
DB 183 TATTAATCCATCTATTCTTGGCCACATCTTCTCTGAGCTACTATTGATTTGGCT 242
QY 1488 TATATTCAAAAGCAATGTGTGCCAAGGAAGAGATTAAAGATGAATATCCATT 1547
DB 243 TATATTCAAAAGCAATGTGTGCCAAGGAAGAGATTAAAGATGAATATCCATT 302
QY 1548 TCCCTGTGCTGCTGTGGGGAATGTGTGCAGAACTACTGCTTCCCTTTTCAAGCACTC 1607
DB 303 TCCCTGTGCTGCTGTGGGGAATGTGTGCAGAACTACTGCTTCCCTTTTCAAGCACTC 362
QY 1608 CTTTGGCCCGAGAGAGAAATGGAAGAACCGAGGAGTGAAGATGCATGCTTCCAGCTG 1667
DB 363 CTTTGGCCCGAGAGAGAAATGGAAGAACCGAGGAGTGAAGATGCATGCTTCCAGCTG 422
QY 1668 TGGCTGTGCTGCGCAGCAAGTCTCATTTGGGGCCAAAGGGGAAACTTTTTTTTGGAGAA 1727
DB 423 TGGCTGTGCTGCGCAGCAAGTCTCATTTGGGGCCAAAGGGGAAACTTTTTTTTGGAGAA 482
QY 1728 GGGGCTTTGCTTTGTCAACCAAGCTGGAATGCAATGCGGGATCTCACTCACCGCAACC 1787
DB 483 GGGGCTTTGCTTTGTCAACCAAGCTGGAATGCAATGCGGGATCTCACTCACCGCAACC 542

```

```

QY 1788 TCACCTCTGGGTTCAAGTGAATTTTCCTGCTCAGCCCTCCCAAGTAGTGGGAATACAG 1847
DB 543 TCACCTCTGGGTTCAAGTGAATTTTCCTGCTCAGCCCTCCCAAGTAGTGGGAATACAG 602
QY 1848 GCACGCCACCATGCGCAGCTAATTTTGTATTTCAGTAGAAACGGGATTTGACCAAGTT 1907
DB 603 GCACGCCACCATGCGCAGCTAATTTTGTATTTCAGTAGAAACGGGATTTGACCAAGTT 662
QY 1908 GGCAGAGCTGTGTCTGAACTCTGACCGCAAGTATCCACCG 1950
DB 663 GGCAGAGCTGTGTCTGAACTCTGACCGCAAGTATCCACCG 705

```

Search completed: November 22, 2003, 12:02:52
Job time : 4745 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2003, 07:52:31 ; Search time 600 Seconds

(without alignments)
10226.371 Million cell updates/sec

Title: US-09-938-803-26

Perfect score: 2273

Sequence: 1 99999gtgaagcgatcgtt.....tttccaaaaa..... 2273

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: N Geneseq_19Jun03.*

```
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2273	100.0	2273	22	AAC88075	Human FLEXHT-6 nuc
2	2211	97.3	2264	21	AA25169	Nucleotide sequenc
3	2186	96.2	2234	22	AA250889	Human receptor-aas
4	1842	81.0	1842	24	ABK84420	Human CDNA differe
5	1842	81.0	1842	24	ABN96918	Human #416 used to
6	1842	81.0	1842	24	ABU64098	Breast cancer rela
7	1461	64.3	1461	22	AA25159	Nucleotide sequenc
8	1156	50.9	1867	23	AA570385	DNA encoding novel

C	9	708	31.1	7461	22	AA530639	DNA encoding novel
C	10	708	31.1	7461	22	AA528701	Genomic sequence #
C	11	708	31.1	7461	25	ACA03402	DNA encoding human
C	12	708	31.1	32174	22	ABA15665	Human nervous syst
C	13	708	31.1	32174	22	ABA19477	Human nervous syst
C	14	708	31.1	32174	22	ABA20359	Human nervous syst
C	15	708	31.1	32174	22	ABA21505	Human nervous syst
C	16	708	31.1	32174	22	ABA36280	Human musculoskele
C	17	708	31.1	32174	22	AA532655	Human genomic DNA
C	18	708	31.1	32174	22	AA534394	Human DNA for a no
C	19	708	31.1	32174	22	AA530638	DNA encoding novel
C	20	708	31.1	32174	22	AA530638	Human reproductive
C	21	708	31.1	32174	22	AA530638	Human reproductive
C	22	708	31.1	32174	22	AA530638	Human reproductive
C	23	708	31.1	32174	22	AA530638	Human reproductive
C	24	708	31.1	32174	22	AA530638	Human reproductive
C	25	708	31.1	32174	22	AA530638	Human reproductive
C	26	708	31.1	32174	22	AA530638	Human reproductive
C	27	708	31.1	32174	22	AA530638	Human reproductive
C	28	708	31.1	32174	22	AA530638	Human reproductive
C	29	708	31.1	32174	22	AA530638	Human reproductive
C	30	708	31.1	32174	22	AA530638	Human reproductive
C	31	708	31.1	32174	22	AA530638	Human reproductive
C	32	708	31.1	32174	22	AA530638	Human reproductive
C	33	708	31.1	32174	22	AA530638	Human reproductive
C	34	708	31.1	32174	22	AA530638	Human reproductive
C	35	708	31.1	32174	22	AA530638	Human reproductive
C	36	708	31.1	32174	22	AA530638	Human reproductive
C	37	708	31.1	32174	22	AA530638	Human reproductive
C	38	708	31.1	32174	22	AA530638	Human reproductive
C	39	708	31.1	32174	22	AA530638	Human reproductive
C	40	708	31.1	32174	22	AA530638	Human reproductive
C	41	708	31.1	32174	22	AA530638	Human reproductive
C	42	708	31.1	32174	22	AA530638	Human reproductive
C	43	708	31.1	32174	22	AA530638	Human reproductive
C	44	708	31.1	32174	22	AA530638	Human reproductive
C	45	596	26.2	1296	23	ABV30301	Human prostate exp

ALIGNMENTS

RESULT 1	
AAC88075	
ID	AAC88075 standard; cDNA; 2273 BP.
AC	AAC88075;
DT	09-MAR-2001 (first entry)
DE	Human FLEXHT-6 nucleotide sequence SEQ ID NO:61.
XX	Human; FLEXHT; full-length molecules expressed in human tissue;
XX	diagnosis; gene expression; genetic linkage; genetic variability;
XX	antianemic; anticonvulsant; antiarteriosclerotic; immunomodulatory;
XX	cytostatic; antiasthmatic; antiinflammatory; hepatotropic; antidiabetic;
XX	anti-gout; antithyroid; neuroprotective; antiarthritic; osteopathic;
XX	anti-psoriatic; antirheumatic; antitumor; gene therapy; anemia; gout;
XX	epilepsy; arteriosclerosis; atherosclerosis; developmental disorder;
XX	cancer; immunological disorder; asthma; bronchitis; cirrhosis;
XX	Crown's disease; diabetes mellitus; Grave's disease; multiple sclerosis;
XX	osteoarthritis; pancreatitis; rheumatoid arthritis; psoriasis;
XX	ulcerative colitis; ss.
OS	Homo sapiens.
XX	
PN	W0200070047-A2.
XX	
PD	23-NOV-2000.
XX	
PF	12-MAY-2000; 2000MO-US13299.
XX	
PR	14-MAY-1999; 99US-0311894.

PR 14-MAY-1999; 99US-0311937.
PR 14-MAY-1999; 99US-0311940.
XX
XX
PA (INCY-) INCYTE GENOMICS INC.
PI Yue H, Tang YT, Lal P, Reddy R, Bacteria S, Baughn MR, Yang J;
PI Azimzai Y, Lu DM, Au-Young J, Shih LL,
XX WPI; 2001-016234/02.
DR P-PSDB; AAB36584.
XX
PT Human FLEXHT protein and DNA sequences, useful for treating
PT immunological disorders, developmental disorders, and cancers -
XX
XX Claim 5; Page 140; 168pp; English.
XX
XX AAC88070 to AAC88124 encode the 55 FLEXHT (full-length molecules
CC expressed in human tissues) proteins given in AAB36579 to AAB36633. The
CC present invention describes an isolated polypeptide (A) comprising an
CC amino acid sequence selected from one of 55 amino acid sequences 42-876
CC residues in length, corresponding to FLEXHT-1 to FLEXHT-55, a 90 %
CC identical sequence, and a biologically active or immunogenic fragment of
CC the sequence. The FLEXHT proteins can have antianaemic, anticoagulant,
CC antiatherosclerotic, immunomodulatory, cytostatic, antiasplenic,
CC antiinflammatory, hepatotropic, antidiabetic, anti-gout, antithyroid,
CC neuroprotective, antilactritic, osteopathic, antipsoriatic, antitumor
CC and antineoplastic activities, and can be used in gene therapy. The
CC polynucleotide sequences can be used to express the protein sequences.
CC Pharmaceutical compositions comprising FLEXHT can be used to treat
CC diseases or conditions associated with altered expression of functional
CC FLEXHT. The proteins and polynucleotides can be used to diagnose and
CC treat disorders including anaemia, epilepsy, arteriosclerosis,
CC atherosclerosis, developmental disorders, cancers, and immunological
CC disorders such as asthma, bronchitis, cirrhosis, Crohn's disease,
CC diabetes mellitus, gout, Grave's disease, multiple sclerosis,
CC osteoarthritis, pancreatitis, psoriasis, rheumatoid arthritis, and
CC ulcerative colitis.
XX
XX Sequence 2273 BP; 494 A; 630 C; 547 G; 602 T; 0 other;
SQ
Query Match 100.0%; Score 2273; DB 22; Length 2273;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGGGTGAAGCATATCTTTGCCCCGATTCGGGGCGCGGACTGGGGGGTCCCTGTG 60
DB 1 GGGGGTGAAGCATATCTTTGCCCCGATTCGGGGCGCGGACTGGGGGGTCCCTGTG 60
QY 61 GGGGCTCCCGAGTTAAGATGCGCTCAGCGGAGGGGAGAGAGGGGACTGTGGTGGCGC 120
DB 61 GGGGCTCCCGAGTTAAGATGCGCTCAGCGGAGGGGAGAGAGGGGACTGTGGTGGCGC 120
QY 121 TGGCGGGGGTCTCAGTCGGGTTTCCAGAGCTGAGCTTAAAGTTGGCGAGTCCC 180
DB 121 TGGCGGGGGTCTCAGTCGGGTTTCCAGAGCTGAGCTTAAAGTTGGCGAGTCCC 180
QY 181 TGGCGCGCTCAGAACAGGCGCTGCGGCTGATCTCTCCATCTTCTGGGTTTACCCTTTTG 240
DB 181 TGGCGCGCTCAGAACAGGCGCTGCGGCTGATCTCTCCATCTTCTGGGTTTACCCTTTTG 240
QY 241 CTTGGTTTATCGGCATTAACCTTTTCTAAGAGAGACCTTACCTCACTCTTCCATA 300
DB 241 CTTGGTTTATCGGCATTAACCTTTTCTAAGAGAGACCTTACCTCACTCTTCCATA 300
QY 301 CTTTACAGGCTCTCATTTGTTTAACTTTTAACTTTGAAACAGCTTACCACTCCCTGC 360
DB 301 CTTTACAGGCTCTCATTTGTTTAACTTTTAACTTTGAAACAGCTTACCACTCCCTGC 360
QY 361 TGTGATTGTGCTTCAAGTTCTCTCATCTTGAAGTGGCGGACCATCACTGCGCTGC 420
DB 361 TGTGATTGTGCTTCAAGTTCTCTCATCTTGAAGTGGCGGACCATCACTGCGCTGC 420
QY 421 TCACTACTTTTGTCTTCCAGATGGCGTCTTCTGGCGTGAATGATTAATCACTGCAACG 480

DB 421 TCACTACTTTTGTCTTCCAGATGGCGTCTTCTGGCGTGAATGATTAATCACTGCAACG 480
QY 481 GCAACTAGCATATCAAGTGAACATGCAATGTGTCTGACTTTGAAGCTGATTTGTT 540
DB 481 GCAACTAGCATATCAAGTGAACATGCAATGTGTCTGACTTTGAAGCTGATTTGTT 540
QY 541 TGGCTGTGACTTACTTTGACGAGGAGAAAGATTCAGAAATTCCTTGTCTCTGAGCAACA 600
DB 541 TGGCTGTGACTTACTTTGACGAGGAGAAAGATTCAGAAATTCCTTGTCTCTGAGCAACA 600
QY 601 AATATGCCATAGCTGTGTCTCTCTGCTGAGAGTTGCTGATTTCTCTTCTATG 660
DB 601 AATATGCCATAGCTGTGTCTCTCTGCTGAGAGTTGCTGATTTCTCTTCTATG 660
QY 661 GGGCTCTTGTGATGGGCCCCAGTTCTCAATGATCAATGATGAGCTGTGCAAGGAG 720
DB 661 GGGCTCTTGTGATGGGCCCCAGTTCTCAATGATCAATGATGAGCTGTGCAAGGAG 720
QY 721 AGCTGATTGACATACAGGAAAGATTCAGAAACAGCATCTCTGCTCTCAAGCGCTGA 780
DB 721 AGCTGATTGACATACAGGAAAGATTCAGAAACAGCATCTCTGCTCTCAAGCGCTGA 780
QY 781 GTCTGGGCTTTTCTAAGTGGCTTACAGACTGTCTGAGCCCCCATACAGAGACT 840
DB 781 GTCTGGGCTTTTCTAAGTGGCTTACAGACTGTCTGAGCCCCCATACAGAGACT 840
QY 841 ATCTCTCACTGAAGCATATGACCAACACCCCTTGTGTTTCCGCTGATGATCATCTGA 900
DB 841 ATCTCTCACTGAAGCATATGACCAACACCCCTTGTGTTTCCGCTGATGATCATCTGA 900
QY 901 TCTGGGCAAGTTGTGCTGTACAAATATGTCACCTGTGCTGCTGACAGAGAGAT 960
DB 901 TCTGGGCAAGTTGTGCTGTACAAATATGTCACCTGTGCTGCTGACAGAGAGAT 960
QY 961 GCAATTTGACGGGCTGAGGCTTCAATGAGCTTTTAAAGAAAGGCAAGCAATGGGATG 1020
DB 961 GCAATTTGACGGGCTGAGGCTTCAATGAGCTTTTAAAGAAAGGCAAGCAATGGGATG 1020
QY 1021 CTTGTGCAACATGAAGTGTGCTTTTGAACAAACCCCGCTTCACTGCAACATTTG 1080
DB 1021 CTTGTGCAACATGAAGTGTGCTTTTGAACAAACCCCGCTTCACTGCAACATTTG 1080
QY 1081 CTTGATTCACATCAACACCAAGCTGAGGCTGCTGATCTTCAATCTTCAAGT 1140
DB 1081 CTTGATTCACATCAACACCAAGCTGAGGCTGCTGATCTTCAATCTTCAAGT 1140
QY 1141 TCTTGGAAATTAAGAACTCTCAGGCTCTCTGTTGATTTCTGAGCCCTGTGGCAGC 1200
DB 1141 TCTTGGAAATTAAGAACTCTCAGGCTCTCTGTTGATTTCTGAGCCCTGTGGCAGC 1200
QY 1201 GCTTGCATCAGGATCTGCTGCTTCCAGATGGAATTCATATTTTATTTGAGAA 1260
DB 1201 GCTTGCATCAGGATCTGCTGCTTCCAGATGGAATTCATATTTTATTTGAGAA 1260
QY 1261 GACAGGCTGCAAGCTCATCAAGAGAGCCCAACCTGAGCAAGCTGGCCGCAATTA 1320
DB 1261 GACAGGCTGCAAGCTCATCAAGAGAGCCCAACCTGAGCAAGCTGGCCGCAATTA 1320
QY 1321 TCTTCCAGCCCTTCTAATTTTGTGTCAGAGCAATCACTGAGCTTCTCATGGGTTACT 1380
DB 1321 TCTTCCAGCCCTTCTAATTTTGTGTCAGAGCAATCACTGAGCTTCTCATGGGTTACT 1380
QY 1381 CCATGACTGCTTGTGCTTCTTCACTGGGCAATAGCTTAAAGTGTATTAATCACTCT 1440
DB 1381 CCATGACTGCTTGTGCTTCTTCACTGGGCAATAGCTTAAAGTGTATTAATCACTCT 1440
QY 1441 AATTCCTTGGCAATCTTCTCTGAGCTTCAATATGCTTATTAATCAAG 1500
DB 1441 AATTCCTTGGCAATCTTCTCTGAGCTTCAATATGCTTATTAATCAAG 1500
QY 1501 CAATGCTGCCAAGAAAGAAAGTAAAGAGATGATATCATTTTCCCTGGTGGCT 1560

QY 407 CATCATGCGGCTCTCACTACCTTTTGTCTCCAGATGGCTTACTCTGTGCTGATACCTA 466
| | | | |
Db 414 CATCATGCGGCTCTCTCACTACCTTTTGTCTCCAGATGGCTTACTCTGTGCTGATACCTA 473
| | | | |
QY 467 TTACACTGCGACGGGCAACTAGATATCAAGTGGAGCAATGCCACATGTGTCTGACTTT 526
| | | | |
Db 474 TTACACTGCGACGGGCAACTAGATATCAAGTGGAGCAATGCCACATGTGTCTGACTTT 533
| | | | |
QY 527 GAACTGATTTGTTTGGCTGTGCTACTCTTTGACGAGGAGAAAGATCAGAAATTCCTTTTC 586
| | | | |
Db 534 GAACTGATTTGTTTGGCTGTGCTACTCTTTGACGAGGAGAAAGATCAGAAATTCCTTTTC 593
| | | | |
QY 587 CTCTGAGCAACAGAAATATGCCATACGTGTGTTCTCTCTGCTGGAAGTGTGCTGTTT 646
| | | | |
Db 594 CTCTGAGCAACAGAAATATGCCATACGTGTGTTCTCTCTGCTGGAAGTGTGCTGTTT 653
| | | | |
QY 647 CTCTACTTCTATGGGGCTTTCTTGTGAGGGGCCAGTTCTCAATGATCACTACATGAA 706
| | | | |
Db 654 CTCTACTTCTATGGGGCTTTCTTGTGAGGGGCCAGTTCTCAATGATCACTACATGAA 713
| | | | |
QY 707 GCTGTGTCAGGAGAGCTGATTTGACATACAGGAAAGATCCAAACAGATCAATTCCTGTC 766
| | | | |
Db 714 GCTGTGTCAGGAGAGCTGATTTGACATACAGGAAAGATCCAAACAGATCAATTCCTGTC 773
| | | | |
QY 767 TCTCAAGCGCTGTAGTCTGGGCTTTTCTACTAGTGGGCTACACACTGCTCAGGCCCCCA 826
| | | | |
Db 774 TCTCAAGCGCTGTAGTCTGGGCTTTTCTACTAGTGGGCTACACACTGCTCAGGCCCCCA 833
| | | | |
QY 827 CATCAGAGAGATCTCTCTCACTGAAAGATATGACAAACCAACCCCTTGTGTTGCTGCTG 886
| | | | |
Db 834 CATCAGAGAGATCTCTCTCACTGAAAGATATGACAAACCAACCCCTTGTGTTGCTGCTG 893
| | | | |
QY 887 CATGTACATGCTGATCTGGGCAAGTTGTGCTGTACAAATATGTCACTGTTGGCTGTG 946
| | | | |
Db 894 CATGTACATGCTGATCTGGGCAAGTTGTGCTGTACAAATATGTCACTGTTGGCTGTG 953
| | | | |
QY 947 CACAGAGAGATATGATTTTGAAGGCGCTGGGCTTCAATGGCTTTGAAGAAAGGCGAA 1006
| | | | |
Db 954 CACAGAGAGATATGATTTTGAAGGCGCTGGGCTTCAATGGCTTTGAAGAAAGGCGAA 1013
| | | | |
QY 1007 GGCAGAGTGGGATGCTGTGCAACATGAAGTGGCTCTTTGAAACAAACCCGCTT 1066
| | | | |
Db 1014 GGCAGAGTGGGATGCTGTGCAACATGAAGTGGCTCTTTGAAACAAACCCGCTT 1073
| | | | |
QY 1067 CACTGGCAGCATTTGCTCAATCAACATCAACCAACGCTGGGTGGCCGCTACATCTT 1126
| | | | |
Db 1074 CACTGGCAGCATTTGCTCAATCAACATCAACCAACGCTGGGTGGCCGCTACATCTT 1133
| | | | |
QY 1127 CAAACGATCAAGTCTCTTGGAAATTAAGAACTCTTCAAGGCTCTCTGTTGCTATTCCT 1186
| | | | |
Db 1134 CAAACGATCAAGTCTCTTGGAAATTAAGAACTCTTCAAGGCTCTCTGTTGCTATTCCT 1193
| | | | |
QY 1187 GGCCCTGTGGCAGGGCTGCACTAGATACCTGCTGTGCTTCCAGATGGAATTCCTCAT 1246
| | | | |
Db 1194 GGCCCTGTGGCAGGGCTGCACTAGATACCTGCTGTGCTTCCAGATGGAATTCCTCAT 1253
| | | | |
QY 1247 TGTATTGTGGAAGAGACAGGCTGCAAGCTCATTTCAAGAGAGCCCAACCTGAGCAAGCT 1306
| | | | |
Db 1254 TGTATTGTGGAAGAGACAGGCTGCAAGCTCATTTCAAGAGAGCCCAACCTGAGCAAGCT 1313
| | | | |
QY 1307 GGCGGCAATTAATGCTCTCCAGCCCTTCTACTATTTTGGTGGAAAGACATCCATGGCT 1366
| | | | |
Db 1314 GGCGGCAATTAATGCTCTCCAGCCCTTCTACTATTTTGGTGGAAAGACATCCATGGCT 1373
| | | | |
QY 1367 CTTGATGGGTTACTCCATGACTGCTTCTGCTCTTCAACGCGGAGCAATGGCTTAAAGT 1426
| | | | |
Db 1374 CTTGATGGGTTACTCCATGACTGCTTCTGCTCTTCAACGCGGAGCAATGGCTTAAAGT 1433
| | | | |
QY 1427 GTATTAATCCATATTTCTTGGCCATCTTCTTCTGAGCCTACTATTCATATTGCC 1486
| | | | |
Db 1434 GTATTAATCCATATTTCTTGGCCATCTTCTTCTGAGCCTACTATTCATATTGCC 1493
| | | | |

QY 1487 TTATATTCAAAAGCAATGCTGCGCAAGGAAAGAGATTAAGAAATGGAATATTCAT 1546
| | | | |
Db 1494 TTATATTCAAAAGCAATGCTGCGCAAGGAAAGAGATTAAGAAATGGAATATTCAT 1553
| | | | |
QY 1547 TTCCCTGTGGGCTGTGCGGGAATGCTGTCAGAAATCTACTGTCCTTTTTCACAGCACT 1606
| | | | |
Db 1554 TTCCCTGTGGGCTGTGCGGGAATGCTGTCAGAAATCTACTGTCCTTTTTCACAGCACT 1613
| | | | |
QY 1607 CCTTTGCCAGAGCAGAGATGGAATGGAAGCCAGGAGGTGGAAGATGATGCTTCCAGCT 1666
| | | | |
Db 1614 CCTTTGCCAGAGCAGAGATGGAATGGAAGCCAGGAGGTGGAAGATGATGCTTCCAGCT 1673
| | | | |
QY 1667 GTGCTCTGCTGCGACCAAGCTTCTGATTTGGGCGCAAGCGGAACTTTTGTGAGA 1726
| | | | |
Db 1674 GTGCTCTGCTGCGACCAAGCTTCTGATTTGGGCGCAAGCGGAACTTTTGTGAGA 1733
| | | | |
QY 1727 AGGCGTCTTGCTTTGTACCCAGCAGTGGATGAGTGGGAGTCTCAGCTCACCGCAAC 1786
| | | | |
Db 1734 AGGCGTCTTGCTTTGTACCCAGCAGTGGATGAGTGGGAGTCTCAGCTCACCGCAAC 1793
| | | | |
QY 1787 CTCACCTCTGTGGGTTCAAGTATTTTCTGCTCAGCCTCCCAAGTACTGGGAATACA 1846
| | | | |
Db 1794 CTCACCTCTGTGGGTTCAAGTATTTTCTGCTCAGCCTCCCAAGTACTGGGAATACA 1853
| | | | |
QY 1847 GGCAGGCCAAGTCCCACTAATTTTGTATTTTCAATGAAACGGGATTTTACCAAGT 1906
| | | | |
Db 1854 GGCAGGCCAAGTCCCACTAATTTTGTATTTTCAATGAAACGGGATTTTACCAAGT 1913
| | | | |
QY 1907 TGGCCAGGCTGTGCTGAACTCCTGACCCGCAAGTGTATCCACCGCTCCGCTCCCAAG 1966
| | | | |
Db 1914 TGGCCAGGCTGTGCTGAACTCCTGACCCGCAAGTGTATCCACCGCTCCGCTCCCAAG 1973
| | | | |
QY 1967 TGTGAGATTAAGGCTGAGCCACCGTCCGCGCCCAAGGGAATCTTGTGGAGG 2026
| | | | |
Db 1974 TGTGAGATTAAGGCTGAGCCACCGTCCGCGCCCAAGGGAATCTTGTGGAGG 2033
| | | | |
QY 2027 AGCAGAGGGGCTCAACATCTCCCTGATTTCCCCCAATGCAATGCTTATCTCCGCA 2086
| | | | |
Db 2034 AGCAGAGGGGCTCAACATCTCCCTGATTTCCCCCAATGCAATGCTTATCTCCGCA 2093
| | | | |
QY 2087 TCTAGCAGGAATCTATTTGTTTCTTCTGCAATTTTACTATGATTTGATGTCGCG 2146
| | | | |
Db 2094 TCTAGCAGGAATCTATTTGTTTCTTCTGCAATTTTACTATGATTTGATGTCGCG 2153
| | | | |
QY 2147 CTACCAACACCCCCCATGAGGGGTGAGAGGGTGCAGAGCCCTGCTCCACTT 2206
| | | | |
Db 2154 CTACCAACACCCCCCATGAGGGGTGAGAGGGTGCAGAGCCCTGCTCCACTT 2213
| | | | |
QY 2207 TTTCTACCTTGAACGTATTAATAAATCACTTCTGTTGTTCAAGTTT 2257
| | | | |
Db 2214 TTTCTACCTTGAACGTATTAATAAATCACTTCTGTTGTTCAAGTTT 2264
| | | | |
RESULT 3
AAZ50889
ID AAZ50889 standard; cDNA; 2234 BP.
XX
AC AAZ50889;
XX
DT 31-MAY-2000 (first entry)
XX
DE Human receptor-associated protein cDNA from Incyte clone 2906971.
XX
XX Human receptor-associated protein; HRAB; Incyte clone 2906971;
KW cytostatic; immunomodulatory; antiinflammatory; cardiant; antianemic;
KW antiarteriosclerotic; hepatotropic; antiallergic antirheumatic;
KW antiaesthetic; osteopathic; antidiabetic; dermatological;
KW neuroprotective; diagnosis; treatment; prevention; reproductive disorder;
KW cardiovascular; cell proliferative; autolimmune; inflammatory; allergy;
KW gastrointestinal; atherosclerosis; cirrhosis; leukaemia; cancer; AIDS;
KW arthritis; anaemia; asthma; dermatitis; diabetes; osteoporosis;
XX multiple sclerosis; irritable bowel syndrome; ss.

OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 63..1526
XX FT /*tag= a
XX FT /product= "HRAP"
XX
XX WO200008155-A2.
XX
XX 17-FEB-2000.
XX
XX
XX 06-AUG-1999; 99WO-US17777.
XX
XX 07-AUG-1998; 98US-0160065.
XX 01-SEP-1998; 98US-0098703.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Hillman JL, Yue H, Lal P, Tang YT, Gorgone GA, Guejler KJ;
XX Corley NC, Baughn MR;
XX
XX MPI: 2000-205710/18.
XX P-PSDB: AAY69987.
XX
XX
XX New human receptor-associated proteins (HRAP) useful for the diagnosis,
XX treatment and prevention of cell proliferative, autoimmune,
XX inflammatory, reproductive, cardiovascular, and gastrointestinal
XX disorders -
XX
XX Claim 9; Pages 89-90; 99pp; English.
XX
XX The present sequence is a cDNA encoding human receptor-associated protein
XX (HRAP) from incyte clone 2906971 obtained from THYMOTOS cDNA library.
XX This sequence is expressed in nervous, gastrointestinal
XX and reproductive tissues. HRAP has cytostatic, immunomodulatory,
XX antiinflammatory, cardiac, antiatherosclerotic, hepatotropic,
XX antithrombotic, antirheumatic, osteopathic, antiallergic, antianaemic,
XX antidiabetic, dermatological and neuroprotective
XX activities. The present sequence is useful in the diagnosis, treatment
XX and prevention of disorders associated with HRAP expression, especially
XX cell proliferative, autoimmune/inflammatory, reproductive,
XX cardiovascular and gastrointestinal disorders (e.g. atherosclerosis,
XX cirrhosis, leukaemia, cancer, AIDS, arthritis, allergies, anaemia,
XX asthma, dermatitis, diabetes, osteoporosis, multiple sclerosis and
XX irritable bowel syndrome).
XX
XX Sequence 2234 BP; 473 A; 626 C; 541 G; 594 T; 0 other;
SQ
Query Match 96.2%; Score 2186; DB 21; Length 2234;
Best local Similarity 100.0%; Pred. No. 0;
Matches 2186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 47 GGGGGGTCCTGTGGGGCTCCGAGTTAAGATGGGCTCTCAGCGAGGGGACAGAGG 106
DB |||||
DB 32 GGGGGGTCCCTGTGGGGCTCCGAGTTAAGATGGGCTCTCAGCGAGGGGACAGAGG 91
QY 107 GACTGTGGTGGCGCTGGCGGGGTTCTGCACTGGGTTCCAGAGAGCTGAAGCTTAACA 166
DB |||||
DB 92 GACTGTGGTGGCGCTGGCGGGGTTCTGCACTGGGTTCCAGAGAGCTGAAGCTTAACA 151
QY 167 GTTGGGACGTCCTGGGCGCTGAGAACAGAGCGCTGGCGCTGATCATCTCATCTTCT 226
DB |||||
DB 152 GTTGGGACGTCCTGGGCGCTGAGAACAGAGCGCTGGCGCTGATCATCTCATCTTCT 211
QY 227 GGGTTACCTTGTGCTTTGTTTATGGGCTTACCTTTTCTACAGAGAGCTGATCTCAT 286
DB |||||
DB 212 GGGTTACCTTGTGCTTTGTTTATGGGCTTACCTTTTCTACAGAGAGCTGATCTCAT 271
QY 287 CCACCTCTTCATACCTTACAGAGCGCTCAATGCTTATTTTAACTTTGAAACAGCT 346
DB |||||
DB 272 CCACCTCTTCATACCTTACAGAGCGCTCAATGCTTATTTTAACTTTGAAACAGCT 331
QY 347 CTACCACTCCCTGTGATGTGCTTCACTTCTATCTTCACTAATGAGCGCGAC 406

DB 332 CTACCACTCCCTGTGATGTGCTTCACTTCTATCTTCACTAATGAGCGCGAC 391
QY 407 CATCACTGGCCCTCACTACCTTTTGTCCAGATGGGCTACCTTGGCTGATACTA 466
DB |||||
DB 392 CATCACTGGCCCTCACTACCTTTTGTCCAGATGGGCTACCTTGGCTGATACTA 451
QY 467 TTACACTGCCACCGCACTACGATATCAAGTGACAATGCCACATTTGTTCTGACTT 526
DB |||||
DB 452 TTACACTGCCACCGCACTACGATATCAAGTGACAATGCCACATTTGTTCTGACTT 511
QY 527 GAACTGATGTGTTGGCTGTGACTACTTTGACGAGGAGAAAGATCAGAAATTCCTGTC 586
DB |||||
DB 512 GAAGCTGATGTGTTGGCTGTGACTACTTTGACGAGGAGAAAGATCAGAAATTCCTGTC 571
QY 587 CTCTGAGCAAGAAATATGCCATACGATGCTGTTCTTCCCTGGGGAAGTTGCTGTTT 646
DB |||||
DB 572 CTCTGAGCAAGAAATATGCCATACGATGCTGTTCTTCCCTGGGGAAGTTGCTGTTT 631
QY 647 CTCTGATCTCTATGGGGCTTCTGTGAGGGCCCGAGTTCTCAATGAATCACTACATGA 706
DB |||||
DB 632 CTCTGATCTCTATGGGGCTTCTGTGAGGGCCCGAGTTCTCAATGAATCACTACATGA 691
QY 707 GCTGTGACGAGAGAGCTGATTTGACATACAGAGAAAGATACCAACAGCATATTCTGTC 766
DB |||||
DB 692 GCTGTGACGAGAGAGCTGATTTGACATACAGAGAAAGATACCAACAGCATATTCTGTC 751
QY 767 TCTCAAGGCTGATGCTGGGCTTTTCTAAGTGGGCTACACATGCTCAGCCCCCA 826
DB |||||
DB 752 TCTCAAGGCTGATGCTGGGCTTTTCTAAGTGGGCTACACATGCTCAGCCCCCA 811
QY 827 CATCAGAGAGCTATCTCTCTCACTGAGACTATGACACCAACCCCTTGTGTCGCTG 886
DB |||||
DB 812 CATCAGAGAGCTATCTCTCTCACTGAGACTATGAGAACCAACCCCTTGTGTCGCTG 871
QY 887 CATGATCATGCTGATCTGGGGAAGTTGTGTGACAAATATGTCACCTGTTGCTGCT 946
DB |||||
DB 872 CATGATCATGCTGATCTGGGGAAGTTGTGTGACAAATATGTCACCTGTTGCTGCT 931
QY 947 CACAGAGAGATGATCTTTTACAGGGCTGGGCTTCAATGCTTTGAAGAAAGGCA 1006
DB |||||
DB 932 CACAGAGAGATGATCTTTTACAGGGCTGGGCTTCAATGCTTTGAAGAAAGGCA 991
QY 1007 GCAAGAGTGGAGTGGCTGGGCAACATGAAGTGGCTTTGAAGCAACCCCGCTT 1066
DB |||||
DB 992 GCAAGAGTGGAGTGGCTGGGCAACATGAAGTGGCTTTGAAGCAACCCCGCTT 1051
QY 1067 CACTGCAACATTTGCTCTATTCATCAATCAACCAACGCTGGTGGCCGCTACATCTT 1126
DB |||||
DB 1052 CACTGCAACATTTGCTCTATTCATCAATCAACCAACGCTGGTGGCCGCTACATCTT 1111
QY 1127 CAAAGACTCAAGTTCTTGGAAATTAAGAACTCTCAGGGTCTCTGCTGATTTCT 1186
DB |||||
DB 1112 CAAAGACTCAAGTTCTTGGAAATTAAGAACTCTCAGGGTCTCTGCTGATTTCT 1171
QY 1187 GGCCTCTGGGACGGGCTGCACTCAGATACCTGGCTCTTCCAGATGGAATTCCTCAT 1246
DB |||||
DB 1172 GGCCTCTGGGACGGGCTGCACTCAGATACCTGGCTCTTCCAGATGGAATTCCTCAT 1231
QY 1247 TGTATTGTGAAAGACAGGCTGCAAGGCTCATTTCAAGAGAGCCCACTGAGCAAGCT 1306
DB |||||
DB 1232 TGTATTGTGAAAGACAGGCTGCAAGGCTCATTTCAAGAGAGCCCACTGAGCAAGCT 1291
QY 1307 GGCCTGATTAATGCTCTCAGGCTTTCACTAATTTGTGCAACAGACATCACTGGCT 1366
DB |||||
DB 1292 GGCCTGATTAATGCTCTCAGGCTTTCACTAATTTGTGCAACAGACATCACTGGCT 1351
QY 1367 CTTCATGGGTTACTGATGAGCTGCTTCTGCTTCAAGTGGGGAACAAATGCTTAAGGT 1426
DB |||||
DB 1352 CTTCATGGGTTACTGATGAGCTGCTTCTGCTTCAAGTGGGGAACAAATGCTTAAGGT 1411
QY 1427 GTATAATCATCTATTTCTTGGCCACATCTTCTTCTGAGCTGATCATATTGCTC 1486
DB |||||

Db 1412 GTATTAATCATCTATTCTTGGCCACATCTTCTCTAGGCTACTATTATTTGCC 1471
 QY 1487 TTATATTACAAAGCAATGGTGGCCAAAGAAAGAAAGTAAAGATGGAATATTCAT 1546
 Db 1472 TTATATTACAAAGCAATGGTGGCCAAAGAAAGAAAGTAAAGATGGAATATTCAT 1531
 QY 1547 TTCCCTGTGTGGCTGTGGGGGAGTGTGCAAGAAATCTGTCTCCCTTTTTCACAGACT 1606
 Db 1532 TTCCCTGTGTGGCTGTGGGGGAGTGTGCAAGAAATCTGTCTCCCTTTTTCACAGACT 1591
 QY 1607 CCTTTGGCCAGACAGAGATGGAAGCAAGGAGGAGTGAAGATGATGCTTCCAGCT 1666
 Db 1592 CCTTTGGCCAGACAGAGATGGAAGCAAGGAGGAGTGAAGATGATGCTTCCAGCT 1651
 QY 1667 GTGCTCTGTGTCGAGCAGCAAGTCTTCAATTTGGGGCCAAAGGGGAAACTTTTGTGAGA 1726
 Db 1652 GTGCTCTGTGTCGAGCAGCAAGTCTTCAATTTGGGGCCAAAGGGGAAACTTTTGTGAGA 1711
 QY 1727 AGGCGTCTGTCTTTGTATCCCAAGCTGTGAATGCAATGCGGGAGTCTAGCTCACCGCAAC 1786
 Db 1712 AGGCGTCTGTCTTTGTATCCCAAGCTGTGAATGCAATGCGGGAGTCTAGCTCACCGCAAC 1771
 QY 1787 CTCACACTCTCGGGTTCAGAGATTTTCTGCTCAGCTCAGCTCCCAAGTAGCTGGGAATACA 1846
 Db 1772 CTCACACTCTCGGGTTCAGAGATTTTCTGCTCAGCTCAGCTCCCAAGTAGCTGGGAATACA 1831
 QY 1847 GGCAGCCACCATGCTCCAGCTAATTTTGTATTTTCAATGAGAAACGGGATTTTCCACAGCT 1906
 Db 1832 GGCAGCCACCATGCTCCAGCTAATTTTGTATTTTCAATGAGAAACGGGATTTTCCACAGCT 1891
 QY 1907 TGGCAGGCTGTGTGTGCAATCTCTGACCCGCAAGATTCACCCGCTCCGCTCCCAAG 1966
 Db 1892 TGGCAGGCTGTGTGTGCAATCTCTGACCCGCAAGATTCACCCGCTCCGCTCCCAAG 1951
 QY 1967 TGGTGGATTAACAGGCGTGAAGCCACGCTGCCCGGCCCAAGGGGAACTCTTGAGGAGG 2026
 Db 1952 TGGTGGATTAACAGGCGTGAAGCCACGCTGCCCGGCCCAAGGGGAACTCTTGAGGAGG 2011
 QY 2027 AGCAGAGGGGCTCAATCTCCCTCTGTATTTCCCATGACATTTGCTTATCTCTCCCA 2086
 Db 2012 AGCAGAGGGGCTCAATCTCCCTCTGTATTTCCCATGACATTTGCTTATCTCTCCCA 2071
 QY 2087 TCTAGCCAGGAATTAATGTGTTTCTTCTGCAATTAATTAATGTAATGTCGCG 2146
 Db 2072 TCTAGCCAGGAATTAATGTGTTTCTTCTGCAATTAATTAATGTAATGTCGCG 2131
 QY 2147 CTACACACACCCCCCATGAGGGGGTGAAGAGGGGCTGCTGCTCCACTT 2206
 Db 2132 CTACACACACCCCCCATGAGGGGGTGAAGAGGGGCTGCTGCTCCACTT 2191
 QY 2207 TTTCTACCTTGAAGCTGTATTAGATA 2232
 Db 2192 TTTCTACCTTGAAGCTGTATTAGATA 2217
 RESULT 4
 ABR84420
 ID ABR84420 standard; cDNA; 1842 BP.
 XX
 AC ABR84420;
 XX
 DT 14-AUG-2002 (first entry)
 XX
 DE Human cDNA differentially expressed in granulocytic cells #991.
 XX
 KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.

XX OS Homo sapiens.
 XX PN NC0200228999-A2.
 XX PD 11-APR-2002.
 XX PF 03-OCT-2001; 2001MO-US30821.
 XX PR 03-OCT-2000; 2000US-237189P.
 XX PA (GENE-) GENE LOGIC INC.
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 DR WPI; 2002-435328/46.
 PT Detecting granulocyte activation by detecting differential expression
 PT of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity
 PS Claim 1; SEQ ID No 991; 114bp; English.
 CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) Gs by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating Gs; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation (especially chronic) in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection, and
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 1842 BP; 413 A; 525 C; 413 G; 491 T; 0 other;
 Query Match 81.0%; Score 1842; DB 24; Length 1842;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 279 TACTATTCACCTCTTCCATTAATTAAGGCTCTCAATGCTTATTTAATTTGA 338
 Db 1 TACTATTCACCTCTTCCATTAATTAAGGCTCTCAATGCTTATTTAATTTGA 60
 QY 339 AACAGCTTACACACTCCCTGTGTATGTGCTTCAATGCTTCAATGCTTCAATG 398

Db 61 AACAGCTTACCACTCCCTGCTGCTGATGCTTCACTTCCTCACTCCCTGACATAATG 120
Qy 399 GGCACGACCATCATCTCCGCTCTCACTACCTTTTGGTTCAGATGGGCTTACCTTGAGCT 458
Db 121 GGCACGACCATCATCTCCGCTCTCACTACCTTTTGGTTCAGATGGGCTTACCTTGAGCT 180
Qy 459 GGATACATATTACCTGCAACCGGCACTACGATATCAAGTATGCAATGCAATGCTATGTT 518
Db 181 GGATACATATTACCTGCAACCGGCACTACGATATCAAGTATGCAATGCAATGCTATGTT 240
Qy 519 CTGACTTTAAGCTGATTTGGTGGCTTTGCTGCTTACCTTTGACGAGGAAAGATCAGAA 578
Db 241 CTGACTTTAAGCTGATTTGGTGGCTTTGCTGCTTACCTTTGACGAGGAAAGATCAGAA 300
Qy 579 TCTTGTCTCTGAGCAAGAAATATGCAATACGATGCTTCTTCCCTGCTGAGAT 638
Db 301 TCTTGTCTCTGAGCAAGAAATATGCAATACGATGCTTCTTCCCTGCTGAGAT 360
Qy 639 GCTGCTTCTCTGCTTCTTCTTATGAGGCTTCTTGGTATGGGCTCCAGTTCTCAATGATC 698
Db 361 GCTGCTTCTCTGCTTCTTCTTATGAGGCTTCTTGGTATGGGCTCCAGTTCTCAATGATC 420
Qy 699 TACATGAACTGCTGAGGAGAGCTGATTTGATACATACGAAAGATACCAACGATC 758
Db 421 TACATGAACTGCTGAGGAGAGCTGATTTGATACATACGAAAGATACCAACGATC 480
Qy 759 ATTCTGCTCTCAAGGCTGAGTCTGAGGCTTTTCTACCTGAGGCTTCAACATGCTC 818
Db 481 ATTCTGCTCTCAAGGCTGAGTCTGAGGCTTTTCTACCTGAGGCTTCAACATGCTC 540
Qy 819 AGCCCCCATCATCAGAAAGATCTCTCTCACTGAAAGCTATGACAAACCCCTTCTG 878
Db 541 AGCCCCCATCATCAGAAAGATCTCTCTCACTGAAAGCTATGACAAACCCCTTCTG 600
Qy 879 TTCCTGCTGATGATCTGCTGATCTGGGCAAGTTTGTCTGTAACAATATGCTACCTG 938
Db 601 TTCCTGCTGATGATCTGCTGATCTGGGCAAGTTTGTCTGTAACAATATGCTACCTG 660
Qy 939 TGGCTGCTGACAGAGAGATGATCTTTGACGGGCTGAGGCTTCAATGGCTTTGAAGA 998
Db 661 TGGCTGCTGACAGAGAGATGATCTTTGACGGGCTGAGGCTTCAATGGCTTTGAAGA 720
Qy 999 AAGGGCAAGCAAGTGGAGTGGCTGCAACATGAAAGTGGCTTTTGAAGCAAA 1058
Db 721 AAGGGCAAGCAAGTGGAGTGGCTGCAACATGAAAGTGGCTTTTGAAGCAAA 780
Qy 1059 CCCCCTTCACTGAGCACTGCTCTATTCACATCAACCAACGCTGGGCTGCGCC 1118
Db 781 CCCCCTTCACTGAGCACTGCTCTATTCACATCAACCAACGCTGGGCTGCGCC 840
Qy 1119 TACATCTTCAAGAGATCTTGAAGTTTGAAGAACTCTTCAGGGCTCTTCGCTG 1178
Db 841 TACATCTTCAAGAGATCTTGAAGTTTGAAGAACTCTTCAGGGCTCTTCGCTG 900
Qy 1179 CTATTCCTGAGCTTGGAGCACTGCACTGAGATCTGCTGCTTCAAGATGAA 1238
Db 901 CTATTCCTGAGCTTGGAGCACTGCACTGAGATCTGCTGCTTCAAGATGAA 960
Qy 1239 TTCCTCATTTGATTTGGAAGACAGGCTGCCAGCTTATTCAGAGAGCCCACTG 1298
Db 961 TTCCTCATTTGATTTGGAAGACAGGCTGCCAGCTTATTCAGAGAGCCCACTG 1020
Qy 1299 AGCAAGCTGGCGGCACTTCTGCTCCAGGCTTCTACTATTTGGTGAACAGACATC 1358
Db 1021 AGCAAGCTGGCGGCACTTCTGCTCCAGGCTTCTACTATTTGGTGAACAGACATC 1080
Qy 1359 CACTGCTCTTCAATGGGTTACTCATGATGCTTCTGCTTCACTGAGCAAAATG 1418
Db 1081 CACTGCTCTTCAATGGGTTACTCATGATGCTTCTGCTTCACTGAGCAAAATG 1140
Qy 1419 CTTAAGGTATTAATCACTATTTCTTGGCAGACATCTTCTCTGAGCTTATTC 1478
Db 1141 CTTAAGGTATTAATCACTATTTCTTGGCAGACATCTTCTCTGAGCTTATTC 1200

Qy 1479 ATATTGCTTATTAATCAAGAAAGCAATGGTGCAGAAAGAAAGAAAGTAAAGATGGA 1538
Db 1201 ATATTGCTTATTAATCAAGAAAGCAATGGTGCAGAAAGAAAGAAAGTAAAGATGGA 1260
Qy 1539 TAATCATTTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1598
Db 1261 TAATCATTTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Qy 1599 ACAGCACTCTTTTGGCCAGAGCAAGAAATGAAAGCCAGGAGCTGAAATGATGATC 1658
Db 1321 ACAGCACTCTTTTGGCCAGAGCAAGAAATGAAAGCCAGGAGCTGAAATGATGATC 1380
Qy 1659 TTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1718
Db 1381 TTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
Qy 1719 TTTGAGAAAGGCTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1778
Db 1441 TTTGAGAAAGGCTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
Qy 1779 ACCGCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1838
Db 1501 ACCGCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
Qy 1839 GGAATACAGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1898
Db 1561 GGAATACAGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
Qy 1899 CACCACTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1958
Db 1621 CACCACTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
Qy 1959 TCCCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2018
Db 1681 TCCCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Qy 2019 GTGGAGGAGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2078
Db 1741 GTGGAGGAGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
Qy 2079 TCTCCCATCTAGCCAGGAATCTATTTGTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 2120
Db 1801 TCTCCCATCTAGCCAGGAATCTATTTGTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1842

RESULT 5
ABN96918
ID ABN96918 standard; DNA; 1842 BP.
XX
AC ABN96918;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #3416 used to diagnose liver cancer.
XX
KM Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
XX metabolic liver tumour; cytostatic; expression profile; disease state;
XX disease progression; drug toxicity; drug efficacy; drug metabolism.
OS Homo sapiens.
XX
PN W0200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
XX (GENE-) GENE LOGIC INC.
PA
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

XX MPI, 2002-426119/45.
 XX Diagnosing and detecting the progression of liver cancer,
 PT hepatocellular carcinoma or metastatic liver tumor in a patient,
 PT involves detecting the level of expression of two or more genes in a
 PT liver tissue sample -
 XX Claim 1; SEQ ID NO 3416; 298bp; English.
 XX The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumor in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatocytic, and
 CC cytostatic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at fwp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 1842 BP; 413 A; 525 C; 413 G; 491 T; 0 other;
 SQ
 Query Match 81.0%; Score 1842; DB 24; Length 1842;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 279 TACCTATCCACCTCTTCCATACCTTTACAGGCTCTCAATTGTTATTTAACTTTGA 338
 DB 1 TACCTATCCACCTCTTCCATACCTTTACAGGCTCTCAATTGTTATTTAACTTTGA 60
 QY 339 AACCACTTACCACTCCCTGCTGTATGTTAGTTCCCTCACTTCCACTATG 398
 DB 61 AACCACTTACCACTCCCTGCTGTATGTTAGTTCCCTCACTTCCACTATG 120
 QY 399 GGGCGACATCACTGCGCTCTCACTACCTTTTGGCTTCAGATGGCTACTTGGCT 458
 DB 121 GGGCGACATCACTGCGCTCTCACTACCTTTTGGCTTCAGATGGCTACTTGGCT 180
 QY 459 GGAATATTAATCACTGCGACCGGCACTAGCATATCAAGTGGACATGCCATTTGTT 518
 DB 181 GGAATATTAATCACTGCGACCGGCACTAGCATATCAAGTGGACATGCCATTTGTT 240
 QY 519 CTGACTTGAAGCTGATTTGGCTTTGGCTTGAATCTTGAACGAGGAGAAAGATCAAGT 578
 DB 241 CTGACTTGAAGCTGATTTGGCTTTGGCTTGAATCTTGAACGAGGAGAAAGATCAAGT 300
 QY 579 TCCCTGTCTCTGAGCAAGAAATATGCATACGTTGTTCCCTTCCCTGCTGGAAGT 638
 DB 301 TCCCTGTCTCTGAGCAAGAAATATGCATACGTTGTTCCCTTCCCTGCTGGAAGT 360
 QY 639 GCTGCTTTCTCTACTTCTATGCGGCTTCTTGGTAGGGCCCACTTCTCATGATCAAC 698
 DB 361 GCTGCTTTCTCTACTTCTATGCGGCTTCTTGGTAGGGCCCACTTCTCATGATCAAC 420
 QY 699 TACATGAAGCTGGGAGGAGAGCTATGATACATCAGGAAAGATACCAAGATC 758
 DB 421 TACATGAAGCTGGGAGGAGAGCTATGATACATCAGGAAAGATACCAAGATC 480
 QY 759 ATTCTCTCTCAAGCCCTGAGTCTGGGCTTTTCTTCACTAGTGGCTACACACTGCTC 818
 DB 481 ATTCTCTCTCAAGCCCTGAGTCTGGGCTTTTCTTCACTAGTGGCTACACACTGCTC 540
 QY 819 AGCCCAATCAAGAAAGTATCTCTCACTGAGAACTATGACAAACCCCTTTGCG 878
 DB 541 AGCCCAATCAAGAAAGTATCTCTCACTGAGAACTATGACAAACCCCTTTGCG 600
 QY 879 TTCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 938

DB 601 TTCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 QY 939 TGGCTGTGACAGAAAGATATGATTTTGAAGGCTGGGCTTCAATGCTTTGAAGA 998
 DB 661 TGGCTGTGACAGAAAGATATGATTTTGAAGGCTGGGCTTCAATGCTTTGAAGA 720
 QY 999 AAGGGCAAGCAAGTGGATGCTGTGCCAATGATGATGATGATGATGATGATGATGAT 1058
 DB 721 AAGGGCAAGCAAGTGGATGCTGTGCCAATGATGATGATGATGATGATGATGATGAT 780
 QY 1059 CCGCGCTTACAGGAGCACTTGGCTTCAATCAATCAATCAATCAATCAATCAATCAATCA 1118
 DB 781 CCGCGCTTACAGGAGCACTTGGCTTCAATCAATCAATCAATCAATCAATCAATCAATCA 840
 QY 1119 TACATCTTCAAGCACTCAAGTTCTTGAATTAAGAACTCTCTGAGGCTCTCTGTT 1178
 DB 841 TACATCTTCAAGCACTCAAGTTCTTGAATTAAGAACTCTCTGAGGCTCTCTGTT 900
 QY 1179 CTATTCCTGGCCCTTGGGACGCGCTGCACTAGATATCTGTTCTTCCAGATGGA 1238
 DB 901 CTATTCCTGGCCCTTGGGACGCGCTGCACTAGATATCTGTTCTTCCAGATGGA 960
 QY 1239 TTCCTATTTGTTATGAGAAAGACAGGCTGCAAGCTCAATCAAGAGAGCCCACTG 1298
 DB 961 TTCCTATTTGTTATGAGAAAGACAGGCTGCAAGCTCAATCAAGAGAGCCCACTG 1020
 QY 1299 AGCAAGCTGGCCCACTTACTGCTCCAGCCCTTCAATTTGGTGCAAGACATC 1358
 DB 1021 AGCAAGCTGGCCCACTTACTGCTCCAGCCCTTCAATTTGGTGCAAGACATC 1080
 QY 1359 CATGCTCTTCAATGGGTTACTGATGATGATGATGATGATGATGATGATGATGATGAT 1418
 DB 1081 CATGCTCTTCAATGGGTTACTGATGATGATGATGATGATGATGATGATGATGATGAT 1140
 QY 1419 CTAAAGTGTATTAATCAATCTATTTCTTGGGCAATCTTCTTCCAGCTACTATTC 1478
 DB 1141 CTAAAGTGTATTAATCAATCTATTTCTTGGGCAATCTTCTTCCAGCTACTATTC 1200
 QY 1479 ATATTCCTTATTAATCAAGAAAGATGTCAGAGAAAGAAAGATTAAGAAAGATGAA 1538
 DB 1201 ATATTCCTTATTAATCAAGAAAGATGTCAGAGAAAGAAAGATTAAGAAAGATGAA 1260
 QY 1539 TAATCATTTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1598
 DB 1261 TAATCATTTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 QY 1599 ACAGCACTCTTGGCCCAAGAGAGAAATGAAAGCCAGGAGGTGAAAGATGATGATG 1658
 DB 1321 ACAGCACTCTTGGCCCAAGAGAGAAATGAAAGCCAGGAGGTGAAAGATGATGATG 1380
 QY 1659 TTCAGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1718
 DB 1381 TTCAGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
 QY 1719 TTTGGAAGAGGCTTTGCTTGTCAACCAAGCTGGAATGCAATGCGGAGATCTCAGCTC 1778
 DB 1441 TTTGGAAGAGGCTTTGCTTGTCAACCAAGCTGGAATGCAATGCGGAGATCTCAGCTC 1500
 QY 1779 ACCGCAACTCCCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1838
 DB 1501 ACCGCAACTCCCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 QY 1839 GGAATACAGGACGCAACATGCGCACTAATTTTGTATTTTCAAGTAAAGGAGATTT 1898
 DB 1561 GGAATACAGGACGCAACATGCGCACTAATTTTGTATTTTCAAGTAAAGGAGATTT 1620
 QY 1899 CACCAAGTTGGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1958
 DB 1621 CACCAAGTTGGCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
 QY 1959 TCCCAAGTGTGGGATTAAGAGCTGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2018

Db 1681 TCCCAAGTCTGGATTACAGCGTGAGCCACCGTCCGCGCCCAAGGGGAACTCTT 1740
QY GTGGAGAGACAGAGGGGCTCACATCTCCCTCTGAATCCCATGACATTTGCCCTTATC 2078
Db 1741 GTGGAGAGACAGAGGGGCTCACATCTCCCTCTGAATCCCATGACATTTGCCCTTATC 1800
QY 2079 TCTCCCATCTAGCCAGGAATCTATTGTGTTTCTTCTGCCC 2120
Db 1801 TCTCCCATCTAGCCAGGAATCTATTGTGTTTCTTCTGCCC 1842

RESULT 6
ABL64098
ID ABL64098 standard; DNA; 1842 BP.
XX
AC ABL64098;
XX
DT 15-MAY-2002 (first entry)
XX
DE Breast cancer related gene sequence SEQ ID NO:2435.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; carcinos;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235638P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236043P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237588P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.

PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
XX (AVALON) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX MPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 2435; 44bp; English.
XX
XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (II)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ Sequence 1842 BP; 413 A; 525 C; 413 G; 491 T; 0 other;
XX

Query Match 81.0%; Score 1842; DB 24; Length 1842;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 TACCTCATCACCCTCTTCATACCTTTACAGGCTCTCAATGCTATTATTTAATTGGA 338
Db 1 TACCTCATCACCCTCTTCATACCTTTACAGGCTCTCAATGCTATTATTTAATTGGA 60

QY 339 AACCACTTACCACTCCTCTGCTGATTTGCTTCACTTCTCAATCTTGACTAATG 398
Db 61 AACCACTTACCACTCCTCTGCTGATTTGCTTCACTTCTCAATCTTGACTAATG 120

QY 399 GGGCGACCATCAGCGCGTCTCTCACTTCTGCTTCCCAATGCGCTACCTTCGGGT 458
Db 121 GGGCGACCATCAGCGCGTCTCTCACTTCTGCTTCCCAATGCGCTACCTTCGGGT 180

QY 459 GGATACCTATTACACTGCGCACCGCAACTACATCAAGTGAATGCAATGCCAATTGTGT 518
Db 181 GGATACCTATTACACTGCGCACCGCAACTACATCAAGTGAATGCAATGCCAATTGTGT 240

QY 519 CTGACTTTGAAGCTGATTTGCTGCTGTTGACTACTTTGACGAGGGAAGATCAAT 578
Db 241 CTGACTTTGAAGCTGATTTGCTGCTGTTGACTACTTTGACGAGGGAAGATCAAT 300

QY 579 TCGTGTCTCTGAGAACAGAAATATGCAATACGAGGTCTTCCCTGCTGGAAGTT 638
Db 301 TCGTGTCTCTGAGAACAGAAATATGCAATACGAGGTCTTCCCTGCTGGAAGTT 360

QY 639 GCTGTTTCTCTTACTTATAGGGCTCTTGTAGGGCCCAAGTTCTCAATGATCAAC 698
Db 361 GCTGTTTCTCTTACTTATAGGGCTCTTGTAGGGCCCAAGTTCTCAATGATCAAC 420

QY 699 TACATGAAGCTGTGAGGAGAGCTGATGACATACCAAGAAATACCAACAGCATC 758
Db 421 TACATGAAGCTGTGAGGAGAGCTGATGACATACCAAGAAATACCAACAGCATC 480

QY 759 ATTCTGCTCTCAAGGCGCTGAGTCTGGGCTTTTCTACTAGTGGGCTACACTGCTC 818
 DB 481 ATTCTGCTCTCAAGGCGCTGAGTCTGGGCTTTTCTACTAGTGGGCTACACTGCTC 540
 QY 819 AGCCCCACATCAAGAGACTATCTCTCACTGAGAAGCTATGACAAACCCCTCTGG 878
 DB 541 AGCCCCACATCAAGAGACTATCTCTCACTGAGAAGCTATGACAAACCCCTCTGG 600
 QY 879 TTCCTGCTGATGATCTGATCTGGGCAAGTTTGTCTGTACAAATATGTCACTGT 938
 DB 601 TTCCTGCTGATGATCTGATCTGGGCAAGTTTGTCTGTACAAATATGTCACTGT 660
 QY 939 TGGCTGTGACAGAGAAGATGATGATTTTGAACGGGCTGGGCTTCAATGGCTTTGAAGA 998
 DB 661 TGGCTGTGACAGAGAAGATGATGATTTTGAACGGGCTGGGCTTCAATGGCTTTGAAGA 720
 QY 999 AAGGGCAAGCAAGTGGAGTGGCTGTGCCAATGAGGTGGCTTTGAACAAC 1058
 DB 721 AAGGGCAAGCAAGTGGAGTGGCTGTGCCAATGAGGTGGCTTTGAACAAC 780
 QY 1059 CCCCGCTTCACTGGACCAATTCCTTCATTCACATCAACCAACGCTGGGCGCCG 1118
 DB 781 CCCCGCTTCACTGGACCAATTCCTTCATTCACATCAACCAACGCTGGGCGCCG 840
 QY 1119 TACATCTTCAAGAGACTCAAGTCTCTTGGAAATTAAGAACTCTCTCAAGGTCTCTGTTG 1178
 DB 841 TACATCTTCAAGAGACTCAAGTCTCTTGGAAATTAAGAACTCTCTCAAGGTCTCTGTTG 900
 QY 1179 CTATTCCTGGCCCTCTGGACGCGCTGACACTAGAGATACCTGCTCTGCTTCCAGATGAA 1238
 DB 901 CTATTCCTGGCCCTCTGGACGCGCTGACACTAGAGATACCTGCTCTGCTTCCAGATGAA 960
 QY 1239 TTCCTCATTTGTTTGTGGAAGACAGGCTGCCAGGCTCATTCAGAGAGAGCCCTG 1298
 DB 961 TTCCTCATTTGTTTGTGGAAGACAGGCTGCCAGGCTCATTCAGAGAGAGCCCTG 1020
 QY 1299 AGCAAGCTGGCGGCAATTCCTTCCTCCAGCCCTTCTACTATTTGGTGGACAGACATC 1358
 DB 1021 AGCAAGCTGGCGGCAATTCCTTCCTCCAGCCCTTCTACTATTTGGTGGACAGACATC 1080
 QY 1359 CACTGCTCTTCATGGGTTACTCCATGACGCTTCCTTCAGTGGAGCAATG 1418
 DB 1081 CACTGCTCTTCATGGGTTACTCCATGACGCTTCCTTCAGTGGAGCAATG 1140
 QY 1419 CTTAAGGTATTAATTCATCTATTTCTTGGCCACATCTTCTGAGCTACTATTC 1478
 DB 1141 CTTAAGGTATTAATTCATCTATTTCTTGGCCACATCTTCTGAGCTACTATTC 1200
 QY 1479 ATATTGCCCTTATTTCAAGAAAGATGGGCCAAGAAAGAAAGTAAAGATGAA 1538
 DB 1201 ATATTGCCCTTATTTCAAGAAAGATGGGCCAAGAAAGAAAGTAAAGATGAA 1260
 QY 1539 TAATCATTTCCCTGGTGGCTGTGGGGAAGTGTGCAAGAACTACTGCTCTCTTTTC 1598
 DB 1261 TAATCATTTCCCTGGTGGCTGTGGGGAAGTGTGCAAGAACTACTGCTCTCTTTTC 1320
 QY 1599 ACAGCACTCTTTTCCCGCAGAGCAAGATGAAAAGCCAGGAGGTGAAGATCATGC 1658
 DB 1321 ACAGCACTCTTTTCCCGCAGAGCAAGATGAAAAGCCAGGAGGTGAAGATCATGC 1380
 QY 1659 TTCACAGTGGCTCTGCTGCGCAGCAAGTCTTATTTGGGGCAAGGGGAACTTTT 1718
 DB 1381 TTCACAGTGGCTCTGCTGCGCAGCAAGTCTTATTTGGGGCAAGGGGAACTTTT 1440
 QY 1719 TTGGAAGAGCGCTTGTGCTTTGTCAACCAAGCTGGAATGCAATGGCGGATCTAGCTC 1778
 DB 1441 TTGGAAGAGCGCTTGTGCTTTGTCAACCAAGCTGGAATGCAATGGCGGATCTAGCTC 1500
 QY 1779 ACCGCAACTCTCACTCTCTGGTTCAAGATATTTCTCTGCTCAGCTCCCAAGATGCTG 1838
 DB 1501 ACCGCAACTCTCACTCTCTGGTTCAAGATATTTCTCTGCTCAGCTCCCAAGATGCTG 1560
 QY 1839 GGAATACAGGCAAGCAACATGCCAGCTAATTTTGTATTTTCAGTAAAGGGAATTT 1898

DB 1561 GGAATACAGGCAAGCGCCATCCAGCTAATTTTGTATTTTCAGTAAACGGGATTT 1620
 QY 1899 CACCAAGTGGGCAAGGCTGTCTCGAATCTCGACCGCAAGTATCCACCGCTCCGC 1958
 DB 1621 CACCAAGTGGGCAAGGCTGTCTCGAATCTCGACCGCAAGTATCCACCGCTCCGC 1680
 QY 1959 TCCCAAGTGTGGGATTACAGGCGTGAGCCACCGTGGCCGCAAGGGGAACTCTT 2018
 DB 1681 TCCCAAGTGTGGGATTACAGGCGTGAGCCACCGTGGCCGCAAGGGGAACTCTT 1740
 QY 2019 GTGGAGAGAGCAAGGCGCTCATCTCCCTCTGAATCCCATGSCACATGGCTTATC 2078
 DB 1741 GTGGAGAGAGCAAGGCGCTCATCTCCCTCTGAATTCCTCATGSCACATGGCTTATC 1800
 QY 2079 TCTCCCATCTAGCCAGAACTAATTTGTTTCTTCTGCC 2120
 DB 1801 TCTCCCATCTAGCCAGAACTAATTTGTTTCTTCTGCC 1842

RESULT 7
 AAF25159
 ID AAF25159 standard; cDNA; 1461 BP.
 XX
 AC AAF25159;
 XX
 DT 30-APR-2001 (first entry)
 XX
 XX
 DE Nucleotide sequence of a human protein having a hydrophobic domain.
 XX
 KW Human; hydrophobic protein; secretory protein; membrane protein; sepsis;
 KW tumour inhibition; immune deficiency; autoimmune disorder; anaemia; burn;
 KW infectious disease; cancer; ulcer; peridontal disease; coagulation;
 KW Parkinson's disease; fertility; immune response; thrombosis; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 1..1461
 FT CDS /tag= a
 FT /product= "hydrophobic protein"
 FT /note= "no termination codon given"
 XX
 XX
 PN WO200104297-A2.
 PD 18-JAN-2001.
 PD 18-JAN-2001.
 PF 16-JUN-2000; 2000MO-JP03942.
 PF 16-JUN-2000; 2000MO-JP03942.
 PR 08-JUL-1999; 99JP-0194359.
 PR 08-JUL-1999; 99JP-0194359.
 XX
 XX (SAGA) SAGAMI CHEM RES CENT.
 PA (PROT-) PROTEGENE INC.
 PA (PROT-) PROTEGENE INC.
 XX
 PI Kato S, Kimura T;
 PI Kato S, Kimura T;
 DR WPI; 2001-103081/11.
 DR WPI; 2001-103081/11.
 DR P-PSDB; AAB31669.
 DR P-PSDB; AAB31669.
 PT Isolated human proteins and polynucleotides are used in research and
 PT have activities including cell proliferation/differentiation activity,
 PT immune stimulating activity and receptor/ligand activity -
 XX
 XX Claim 3; Page 108-109; 151p; English.
 XX
 XX The present sequence encodes a human protein with hydrophobic domains.
 CC AAF25159 represents a longer version of the present sequence. The
 CC protein possesses a hydrophobic domain and so is a secretory protein
 CC or a membrane protein. The protein is used as an antigen to prepare
 CC antibodies. The polynucleotide sequence is useful as a source of probes
 CC for genetic diagnosis. It is also useful for producing the protein
 CC in large quantities and for gene therapy. The eukaryotic cells are used
 CC for detecting the receptors or ligands corresponding to the protein and

for detecting small novel pharmaceuticals. The antibodies are also used for detection, quantification and purification of the proteins. Both the protein and polynucleotide may be used in research or as nutritional sources or supplements. The protein may have cytokine and cell proliferation/differentiation activity, immune stimulating or suppressing activity, hematopoietic regulating activity, tissue growth activity, activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity and tumour inhibition activity. It may therefore may be used to treat immune deficiencies resulting from autoimmune disorders or infectious diseases, cancer, sepsis, anaemias, burns and ulcers, periodontal disease, Parkinson's disease, induce fertility, improve immune response and enhance coagulation or inhibit thrombolysis.

SQ Sequence 1461 BP; 326 A; 401 C; 336 G; 398 T; 0 other;

Query Match 64.3%; Score 1461; DB 22; Length 1461;

Best Local Similarity 100.0%; Pred. No. 0; Matches 1461; Contaminating 0; Misclassified

Matches 1461; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	78	ATGGGCGTCTCTCAGCGGAGGGGACAGAGGGGACCTGTGTGGCGCTGGCGGGGGCTTCGACG	137
Dp	1	ATGGCGTCTCTCAGCGGAGGGGACAGAGGGGACCTGTGTGGCGCTGGCGGGGGTTCGACG	60
QY	138	TCGGGTTTCCAGAGCTGAGCCTTAACAAGTTGGCAGCGTCCCTGGGCGGCTCAGAACG	197
Dp	61	TCGGGTTTCCAGAGCTGAGCCTTAACAAGTTGGCGACGTCCTGGGCGGCTCAGAACG	120
QY	198	GGCGTCGGGCTGATCATCTCCATCTTCTCTGGGTACCCCTTTGCTTTGTTTATGGGCAT	257
Dp	121	GGCGTCGGGCTGATCATCTCCATCTTCTCTGGGTACCCCTTTGCTTTGTTTATGGGCAT	180
QY	258	TACCTTTTCTACAGAGAGACTTACCTCATCCACCTTTCGCATTCACCTTTCAGGCGCTCTCA	317
Dp	181	TACCTTTTCTACAGAGAGACTTACCTCATCCACCTTTCGCATTCACCTTTCAGGCGCTCTCA	240
QY	318	ATTGCTTATTTTAACTTTTGAAAACCAAGCTCTACACACTCCGCTGTGTATATTGTCTTCAG	377
Dp	241	ATTGCTTATTTTAACTTTTGAAAACCAAGCTCTACACACTCCGCTGTGTATATTGTCTTCAG	300
QY	378	TTCTTCATTCCTTCGACTATATGGGCGGCACATCATCTGCCGTCTCTCATCTTTTGCTTC	437
Dp	301	TTCTTCATTCCTTCGACTATATGGGCGGCACATCATCGCGTCTCTCATCTTTTGCTTC	360
QY	438	CAGATGGCGCTTACCTTCGCGTGGATCTATTTACACTGCGCACCGGCAACTACATATCAG	497
Dp	361	CAGATGGCGCTTACCTTCGCGTGGATCTATTTACACTGCGCACCGGCAACTACATATCAG	420
QY	498	TGGACAATGCCACATTGTGTCTGACTTTTGAAGCTAATTGTTGGCTGTGACTACTTT	557
Dp	421	TGGACAATGCCACATTGTGTCTGACTTTTGAAGCTAATTGTTGGCTGTGACTACTTT	480
QY	558	GACGGAGGGAAGATCAGAAATTCCTGTCTCTGAGCAACAGAAATATGCGATACGTGT	617
Dp	481	GACGGAGGGAAGATCAGAAATTCCTGTCTCTGAGCAACAGAAATATGCGATACGTGT	540
QY	618	GTTCCCTCCCTCTGGAAGTTGCTGTCTCTCTCACTTCTATGAGGCGCTTCTGTGTAGGG	677
Dp	541	GTTCCCTCCCTCTGGAAGTTGCTGTCTCTCTCACTTCTATGAGGCGCTTCTGTGTAGGG	600
QY	678	CCCCAGTTTCTCAATGAATCACTACATGAAAGCTGTGTACAGGAGAGCTGATTTGACATCCA	737
Dp	601	CCCCAGTTTCTCAATGAATCACTACATGAAAGCTGTGTACAGGAGAGCTGATTTGACATCCA	660
QY	738	GGAAGATATCCAAAGACATCAATTCCTGTCTCAAGCGCTGAGTCTGGGCGCTTTTCTTAC	797
Dp	661	GGAAGATATCCAAAGACATCAATTCCTGTCTCTCAAGCGCTGAGTCTGGGCGCTTTTCTTAC	720
QY	798	CTAGTGGGCTACACACTGCTCAAGCCCCCACAATCAGAGAATATCTCTCCATCTGAAGAC	857
Dp	721	CTAGTGGGCTACACACTGCTCAAGCCCCCACAATCAGAGAATATCTCTCCATCTGAAGAC	780
QY	858	TATGACAAACACCCCTTCTGTGTTCCGCTGCATGTACATGCTGATCTGGGGCAAGTTTGTG	917

[illegible]

XX Drmanac RT, Liu C, Tang YT;
PI MPI; 2001-639362/73.
XX P-PSDB; ABG06198.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
PS Claim 1; SEQ ID No 6189; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. A564197-A594564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1867 BP; 421 A; 532 C; 433 G; 481 T; 0 other;
Query Match 50.9%; Score 1156; DB 23; Length 1867;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1716; Conservative 0; Mismatches 4; Indels 3; Gaps 3;
QY 266 CTACAGAGAGACCTACCTCATCCCTCTTCATCTTACAGAGCTCTCATTTGGTTA 325
DB 1 CTACAGAGAGACCTACCTCATCCCTCTTCATCTTACAGAGCTCTCATTTGGTTA 60
QY 326 TTTTAACTTTGAAACCAAGCTTACCACTCCCTGCTGTATGTGCTTCACTTCAT 385
DB 61 TTTTAACTTTGAAACCAAGCTTACCACTCCCTGCTGTATGTGCTTCACTTCAT 120
QY 386 CCTTCACATAATGGCCGCGACCATCACTGCGCTCTTACATCTTTTGTCCAGATGCG 445
DB 121 CCTTCGACTATAGGGCGCGACCATCACTGCGCTCTTACATCTTTTGTCCAGATGCG 180
QY 446 CTACCTCTGGCTGATATATTACACTGCGACCGGCACTACGATCACTCAAGTGCACAT 505
DB 181 CTACCTCTGGCTGATATATTACACTGCGACCGGCACTACGATCACTCAAGTGCACAT 240
QY 506 GCCACATTTGTCTGACTTTGAAGCTGATTTGGTGTGACTGACTTTTGAACGAGG 565
DB 241 GCCACATTTGTCTGACTTTGAAGCTGATTTGGTGTGACTGACTTTTGAACGAGG 300
QY 566 GAAAGATCAGAAATCTTGTCTGCTTGAGCAACAGAAATATAGCCATAGTGTGTTCTTC 625
DB 301 GAAAGATCAGAAATCTTGTCTGCTTGAGCAACAGAAATATAGCCATAGTGTGTTCTTC 360
QY 626 CCTGCTGAGAGTGTGCTGTTCTCTTACTTATAGGGGCTTCTTGTGAGGGCCCAAGT 685
DB 361 CCTGCTGAGAGTGTGCTGTTCTCTTACTTATAGGGGCTTCTTGTGAGGGCCCAAGT 420
QY 686 CTGAATGAATCATTACATGAAGCTGTGCGAGGAGAGCTGATTTGACATACAGAAAGAT 745
DB 421 CTGAATGAATCATTACATGAAGCTGTGCGAGGAGAGCTGATTTGACATACAGAAAGAT 480

QY 746 ACCAAACAGATCATTTCTGCTCTCAAGGCGCTGAGTCTGGGCTTTTCTACCTAGTGG 805
DB 481 ACCAAACAGATCATTTCTGCTCTCAAGGCGCTGAGTCTGGGCTTTTCTACCTAGTGG 540
QY 806 CTACACACTGTGAGCCGCCACATCAGAGAAGCTATCTCTCACTGAGACTATGACAA 865
DB 541 CTACACACTGTGAGCCGCCACATCAGAGAAGCTATCTCTCACTGAGACTATGACAA 600
QY 866 CCACCCCTTCTGTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 925
DB 601 CCACCCCTTCTGTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 926 ATATGTCACCTGTTGGCTGTGATCAGAGAAGATATGATATTTGACGGGCTGGCTTCA 985
DB 661 ATATGTCACCTGTTGGCTGTGATCAGAGAAGATATGATATTTGACGGGCTGGCTTCA 720
QY 986 TGGCTTTGAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1045
DB 721 TGGCTTTGAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780
QY 1046 CTTTGAACAAACCCCGCTTCACTGGACCAATGGCTGATTCATCAATCAACCAACGC 1105
DB 781 CTTTGAACAAACCCCGCTTCACTGGACCAATGGCTGATTCATCAATCAACCAACGC 840
QY 1106 CTGGGTGGCCCGCTATATCTTCAAAAGACTCAAGTTCCTGAAATTAAGAACTCTCTCA 1165
DB 841 CTGGGTGGCCCGCTATATCTTCAAAAGACTCAAGTTCCTGAAATTAAGAACTCTCTCA 900
QY 1166 GGGTCTCTC-GTGTCTATTCTGAGCCCTGAGCAGGCTGACACTGAGATACCTGATC 1224
DB 901 AGGTCTCTGAGTGTGATCTCTGCTCTGAGCAGGCTGACACTGAGATACCTGATC 960
QY 1225 GCTTCAGATGATATCTCTATTTGTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1284
DB 961 GCTTCAGATGATATCTCTATTTGTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1020
QY 1285 AGAGCCCAACCCGAGACAGCTGGCGGCAATTAAGTCTCCAGCCCTTCTATTTTGG 1344
DB 1021 AGAGCCCAACCCGAGACAGCTGGCGGCAATTAAGTCTCCAGCCCTTCTATTTTGG 1080
QY 1345 TGCACAGACATCCATCTGCTCTTATGAGGTTACTCAATGATGCTTCTGCTCTTCA 1404
DB 1081 TGCACAGACATCCATCTGCTCTTATGAGGTTACTCAATGATGCTTCTGCTCTTCA 1140
QY 1405 CGT-GGACAAATGCTTAAAGTGTATTAATCACTATTTCTTGGCAATCTTCTTC 1463
DB 1141 CGTGGGGAACAAATGCTTAAAGTGTATTAATCACTATTTCTTGGCAATCTTCTTC 1200
QY 1464 CTGAGCTCTATTCATTTGCTTATTTGCTTATTTGCTTATTTGCTTATTTGCTTATTT 1522
DB 1201 CTGAGCTCTATTCATTTGCTTATTTGCTTATTTGCTTATTTGCTTATTTGCTTATTT 1260
QY 1523 GTTAAAGAAATGAAATTAATTCATTTCCCTGTTGGCTGTGCGGACTGGTGAGAACT 1582
DB 1261 GTTAAAGAAATGAAATTAATTCATTTCCCTGTTGGCTGTGCGGACTGGTGAGAACT 1320
QY 1583 ACTGTTCTCTTTTCAACAGCACTCTTTGCCCCAGAGAGAGATGAGAAACCAAGGGA 1642
DB 1321 ACTGTTCTCTTTTCAACAGCACTCTTTGCCCCAGAGAGAGATGAGAAACCAAGGGA 1380
QY 1643 GGTGAAGATCAGATGCTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1702
DB 1381 GGTGAAGATCAGATGCTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
QY 1703 AAAGGGAAATCTTTTGTGAGAAAGGCTTTGCTTTGTCACCAAGCTGAGATGAGT 1762
DB 1441 AAAGGGAAATCTTTTGTGAGAAAGGCTTTGCTTTGTCACCAAGCTGAGATGAGT 1500
QY 1763 GGGGGATCTCAGCTACCGCAACTCTCAGCTCTGAGGTTCAAGTATTTTCTGCTCA 1822
DB 1501 GGGGGATCTCAGCTACCGCAACTCTCAGCTCTGAGGTTCAAGTATTTTCTGCTCA 1560
QY 1823 GCTCCCAAGTACTGGGAATACAGGACAGCCCAATGCCCACTAATTTTGTATTTTC 1882

Db 1561 GCCTCCCAAGTAGCTGGAAATACAGSCAGCCAGCCAGTCCAGCTATTTTGTATTTTC 1620
 Qy 1883 AATTAGAAAGGATTTTACCAGCTTGGCCAGGCTGTCTCGAATCTCTACCGCAAGTA 1942
 Db 1621 AATTAGAAAGGATTTTACCAGCTTGGCCAGGCTGTCTCGAATCTCTACCGCAAGTA 1680
 Qy 1943 TCCACCCGCTCCGCTCCCAAGTCTGGATTACAGGCGTG 1985
 Db 1681 TCCACCCGCTCCGCTCCCAAGTCTGGATTACAGGCGTG 1723

RESULT 9
 AAS30639/C
 ID AAS30639 standard; DNA; 7461 BP.
 XX AAS30639;
 XX 21-NOV-2001 (first entry)
 XX
 DE DNA encoding novel lung cancer antigen, Seq ID No 91.
 XX
 XX Human; lung cancer; immunosuppressive; antiarthritic; antirheumatic;
 XX antiproliferative; cytostatic; cardiatic; vasotropic; cerebroprotective;
 XX nocotropic; neuroprotective; antibacterial; virucide; fungicide;
 XX ophthalmological; autoimmune disease; rheumatoid arthritis; neoplasm;
 XX hyperproliferative disorder; cardiovascular disorder; angiosgenesis;
 XX nervous system disorder; Alzheimer's disease; infection; skin aging;
 XX ocular disorder; wound healing; organ transplantation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200155300-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01238.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR MPI; 2001-465565/50.
 XX
 PT Isolated nucleic acid molecule encoding a lung cancer antigen is used
 XX in preventing, treating or ameliorating a medical condition
 XX
 PS Disclosure: SEQ ID No 91; 475bp; English.
 XX
 CC The invention relates to novel isolated lung cancer antigen
 CC polynucleotides (I) and polypeptides (II). (I) and (II) are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, cats, dogs, chickens or sheep. (I) and (II) are
 CC also used in diagnosing a pathological condition or susceptibility to a
 CC pathological condition, in particular lung cancer. The antibodies to
 CC (II) can also be used in alleviating symptoms associated with the
 CC disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme
 CC linked immunosorbent assays (ELISA). Disorders which are diagnosed or
 CC treated include autoimmune diseases e.g. Rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
 CC ocular disorders e.g. corneal infection. The polypeptides can also be
 CC used to aid wound healing and epithelial cell proliferation, to prevent
 CC skin aging due to sunburn, to maintain organs before transplantation, for
 CC supporting cell culture of primary tissues, to regenerate tissues and in
 CC chemotaxis. AAS30580-AAS30655 represent novel human lung cancer antigen
 CC coding sequences, PCR primers and related sequences of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly

CC from WIFO at: ftp.wifo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 7461 BP; 1883 A; 1666 C; 1923 G; 1989 T; 0 other;
 Query Match 31.1%; Score 708; DB 22; Length 7461;
 Best Local Similarity 100.0%; Pred. No. 3e-300;
 Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1553 GGTGGCTGTGGGAGCTGGGAGCAAACTCTGCTCCCTTTACAGCACTCTG 1612
 Db 714 GGTGGCTGTGGGAGCTGGGAGCAAACTCTGCTCCCTTTACAGCACTCTG 655
 Qy 1613 CCCAGAGCAGAGATGAAAAAGCCAGGAGGTGAGATCGATCTTCAGCTGTGCT 1672
 Db 654 CCCAGAGCAGAGATGAAAAAGCCAGGAGGTGAGATCGATCTTCAGCTGTGCT 595
 Qy 1673 CTGCTGCCAGCCCACTTTCAATTGGGGCCAAAGGGGAACTTTTGTGAGAAAGCGT 1732
 Db 594 CTGCTGCCAGCCCACTTTCAATTGGGGCCAAAGGGGAACTTTTGTGAGAAAGCGT 535
 Qy 1733 CTGCTTTGTATCACCAGCTGGAAATGACATGGCGGATCTCAGCTCACCAGCACTCCAC 1792
 Db 534 CTGCTTTGTATCACCAGCTGGAAATGACATGGCGGATCTCAGCTCACCAGCACTCCAC 475
 Qy 1793 CTGCTGGGTTCAAGGATTTTCTGCTCAGCTCCCAAGTGTGGGAAATACAGCAGC 1852
 Db 474 CTGCTGGGTTCAAGGATTTTCTGCTCAGCTCCCAAGTGTGGGAAATACAGCAGC 415
 Qy 1853 CCACCATGCCCACTAATTTTGTATTTTCAATAGAAACGGGATTTACACAGTTGGCCA 1912
 Db 414 CCACCATGCCCACTAATTTTGTATTTTCAATAGAAACGGGATTTACACAGTTGGCCA 355
 Qy 1913 GGTGCTGTCCGAATCTTCGACCGCAAGTATCCACCCGCTCCGCTCCCAAGTGTGG 1972
 Db 354 GGTGCTGTCCGAATCTTCGACCGCAAGTATCCACCCGCTCCGCTCCCAAGTGTGG 295
 Qy 1973 GATTACAGGCGTGAACCCGCGCCGCGCCCAAGGAGAACTTTGTGGAGAGCAGA 2032
 Db 294 GATTACAGGCGTGAACCCGCGCCGCGCCCAAGGAGAACTTTGTGGAGAGCAGA 235
 Qy 2033 GGGGCTCAATCTCCCTCTGATTTCCCATGACATTCCTTATCTCTCCCATCTAGC 2092
 Db 234 GGGGCTCAATCTCCCTCTGATTTCCCATGACATTCCTTATCTCTCCCATCTAGC 175
 Qy 2093 CAGGAATCTATGCTTTTCTTCTGCAATTAATGATGATGATGCGGCTACCA 2152
 Db 174 CAGGAATCTATGCTTTTCTTCTGCAATTAATGATGATGATGCGGCTACCA 115
 Qy 2153 CCACCCCGCCCATGGGGGGGTGAGAGGGGTGCAAGCCCTGCTCCACTTTTCTTA 2212
 Db 114 CCACCCCGCCCATGGGGGGGTGAGAGGGGTGCAAGCCCTGCTCCACTTTTCTTA 55

RESULT 10
 AAS28701/C
 ID AAS28701 standard; DNA; 7461 BP.
 XX AAS28701;
 XX 07-NOV-2001 (first entry)
 XX
 DE Genomic sequence #541 encoding for novel human respiratory antigen.
 XX
 XX Human; respiratory antigen; respiratory disorder; throat disorder;
 XX lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;
 XX anti allergic; anti asthmatic; anti inflammatory; oilyfactory;
 XX respiratory active; ds.
 OS Homo sapiens.

PR 21-SEP-2000; 2000US-234274P.
 PR 25-SEP-2000; 2000US-234997P.
 PR 27-SEP-2000; 2000US-235834P.
 PR 29-SEP-2000; 2000US-236327P.
 PR 29-SEP-2000; 2000US-236367P.
 PR 29-SEP-2000; 2000US-236368P.
 PR 29-SEP-2000; 2000US-236369P.
 PR 29-SEP-2000; 2000US-236370P.
 PR 02-OCT-2000; 2000US-236802P.
 PR 02-OCT-2000; 2000US-237037P.
 PR 02-OCT-2000; 2000US-237038P.
 PR 02-OCT-2000; 2000US-237040P.
 PR 13-OCT-2000; 2000US-239935P.
 PR 20-OCT-2000; 2000US-240960P.
 PR 20-OCT-2000; 2000US-241785P.
 PR 20-OCT-2000; 2000US-241809P.
 PR 01-NOV-2000; 2000US-244617P.
 PR 17-NOV-2000; 2000US-249299P.
 PR 08-DEC-2000; 2000US-251856P.
 PR 08-DEC-2000; 2000US-251868P.
 PR 08-DEC-2000; 2000US-251869P.
 XX
 PA (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 DR MPI; 2003-311001/30.
 XX
 PT Novel lung cancer related polypeptide useful for detecting, treating,
 PT preventing and/or prognosing lung, respiratory, hyperproliferative,
 PT immune system, cardiovascular endocrine or gastrointestinal disorders
 XX
 XX Disclosure; Page 282-286; 332pp; English.
 XX
 CC The invention relates to an isolated lung cancer related polypeptide its
 CC variants, allelic variants or homologues. The polypeptide and its nucleic
 CC acid are useful for preventing, treating, or ameliorating a medical
 CC condition in a mammalian subject, for diagnosing a pathological condition
 CC or a susceptibility to a pathological condition in a subject by
 CC determining the presence or absence of mutation in the nucleic acid or
 CC biological sample and diagnosing a pathological condition based on the
 CC result. The polypeptide, antibodies to the polypeptide or the
 CC polynucleotide are useful for detecting, treating, preventing and/or
 CC prognosing disorders of the lung, such as small cell lung cancer, non-
 CC small cell lung cancer (e.g. squamous cell carcinoma, adenocarcinoma,
 CC large cell carcinoma, adenocarcinoma and undifferentiated
 CC carcinoma) or lung cancer metastasis. The polypeptide, its antibodies or
 CC its polynucleotide are also useful for detecting, treating, preventing
 CC and/or prognosing respiratory disorders e.g. nonallergic rhinitis and
 CC adult respiratory distress syndrome; hyperproliferative disorders e.g.
 CC lung neoplasms and prostate neoplasms; immune system disorders e.g.
 CC rheumatoid arthritis and multiple sclerosis; blood-related disorders e.g.
 CC anaemia and leukopenia; allergic reactions e.g. anaphylaxis and asthma;
 CC inflammatory disorders e.g. appendicitis and dermatitis; graft vs host
 CC disease; immune complex disease e.g. serum sickness and polyarteritis
 CC nodosa; urinary system disorders e.g. glomerulonephritis and kidney
 CC failure; cardiovascular disorders e.g. arrhythmia and myocardial
 CC infarction; musculoskeletal system disorders e.g. Albers-Schönberg
 CC disease and osteoarthritis; neurological disorders e.g. Alzheimer's
 CC disease and Parkinson's disease; endocrine disorders e.g. Addison's
 CC disease and diabetes mellitus; gastrointestinal disorders e.g. ulcers and
 CC gastric reflux; reproductive system disorders e.g. infertility and
 CC impotence; developmental and inherited disorders e.g. Fanconi's syndrome
 CC and cerebral palsy; diseases at the cellular level e.g. cancer and
 CC amyotrophic lateral sclerosis; infectious diseases e.g. viral and
 CC bacterial. The polypeptide, its antibodies or its polynucleotide are also
 CC useful for wound healing e.g. grafts and burns and for regeneration of
 CC tissues damaged due to e.g. aging and trauma. The present sequence

CC represents DNA encoding a human lung cancer antigen.
 XX
 XX Sequence 7461 BP, 1883 A, 1666 C, 1923 G, 1969 T; 0 other:
 SQ
 Query Match 31.1%; Score 708; DB 25; Length 7461;
 Best Local Similarity 100.0%; Pred. No. 3e-300;
 Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1553 GGTGGCTTGTGGGAGCTGTGTGACAAATCTATCTCTCTTTTACAGCACTCTTTG 1612
 DB 714 GGTGGCTTGTGGGAGCTGTGTGACAAATCTATCTCTCTTTTACAGCACTCTTTG 655
 QY 1613 CCCGAGAGAGAAAGAAAGCCGAGAGGTGAAGATGATGCTTCAGCTGTGCTT 1672
 DB 654 CCCGAGAGAGAAAGAAAGCCGAGAGGTGAAGATGATGCTTCAGCTGTGCTT 595
 QY 1673 CTGCTCCAGCCAGCAAGCTTTCATTTGGGGCCAAAGGGGAACTTTTGGAGAGCGCT 1732
 DB 594 CTGCTCCAGCCAGCAAGCTTTCATTTGGGGCCAAAGGGGAACTTTTGGAGAGCGCT 535
 QY 1733 CTGCTTGTGTACCCAGCTGTGAATGATGCTGTGGGATCTCAGCTACCGCACTTCAC 1792
 DB 534 CTGCTTGTGTACCCAGCTGTGAATGATGCTGTGGGATCTCAGCTACCGCACTTCAC 475
 QY 1793 CTGCTGGGTTCAGATATTTCTGCTCAGCTCCCAAGTACGAGATACAGGACAG 1852
 DB 474 CTGCTGGGTTCAGATATTTCTGCTCAGCTCCCAAGTACGAGATACAGGACAG 415
 QY 1853 CCACATGCCCAGCTAATTTTGTATTTTCAGTAGAAAGGGAATTCACCACTGTGGCCA 1912
 DB 414 CCACATGCCCAGCTAATTTTGTATTTTCAGTAGAAAGGGAATTCACCACTGTGGCCA 355
 QY 1913 GGTGTGTTCGAATCTCTGACCGCAAGTATCCACCCGCTCCGCTCCCAAGTGTGG 1972
 DB 354 GGTGTGTTCGAATCTCTGACCGCAAGTATCCACCCGCTCCGCTCCCAAGTGTGG 295
 QY 1973 GATTACAGGGGTGAGCACCGGTCCGCGCCCAAGGGGAACTCTGTGGAGAGAGAGA 2032
 DB 294 GATTACAGGGGTGAGCACCGGTCCGCGCCCAAGGGGAACTCTGTGGAGAGAGAGA 235
 QY 2033 GGGGCTCACATCTCCCTGTGATTCGCCCATGACATGCTTATCTCTCCCATTAAGC 2092
 DB 234 GGGGCTCACATCTCCCTGTGATTCGCCCATGACATGCTTATCTCTCCCATTAAGC 175
 QY 2093 CAGGAATCTATGTGTGTCTTCTGTGCAATTTACTATGATTTGTATGTGCGCTACCA 2152
 DB 174 CAGGAATCTATGTGTGTCTTCTGTGCAATTTACTATGATTTGTATGTGCGCTACCA 115
 QY 2153 CACCCCCCCCCATGGGGGGGTGAGAGGGGTGCAGAGCCCTGCTCTCCACTTTTCTA 2212
 DB 114 CACCCCCCCCCATGGGGGGGTGAGAGGGGTGCAGAGCCCTGCTCTCCACTTTTCTA 55
 QY 2213 CCTTGAACGTATTAATTAATCACTTCTGTGTTCAGTTTTC 2260
 DB 54 CCTTGAACGTATTAATTAATCACTTCTGTGTTCAGTTTTC 7
 RESULT 12
 ABA1565/c
 ID ABA1565 standard; DNA; 32174 BP.
 XX
 AC ABA1565;
 XX
 DT 23-JAN-2002 (first entry)
 XX
 DE Human nervous system related polynucleotide SEQ ID NO 7996.
 XX
 KW Human, noctropic; neuroprotective; cytoskeletal; dermatological; vitruide;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
 KW antiparkinsonian; antischilling; antianaemic; antiallergic; cancer;
 KW antitubercular; hepatotropic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antileptic; anticonvulsant; antifungal;
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;

XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 7996; 1701bp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABAI1004-ABA21534) and proteins
CC (ABAI4678-ABAI1801) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 32174 BP; 8001 A; 6837 C; 7064 G; 8747 T; 1525 other;
Query Match 31.1%; Score 708; DB 22; Length 32174;
Best Local Similarity 100.0%; Pred. No. 2.8e-300;
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1553 GGTGGCTGTGGGGAGTGTGCAAGAACTACTGCTCTCCCTTTTCACAGACTCTTTG 1612
DB 6043 GGTGGCTGTGGGGAGTGTGCAAGAACTACTGCTCTCCCTTTTCACAGACTCTTTG 5984
QY 1613 CCCAGAGCAGAGATGGAAGAAAGCAGAGAGTGAAGTCAAGCTTCAGCTTCCT 1672
DB 5983 CCCAGAGCAGAGATGGAAGAAAGCAGAGAGTGAAGTCAAGCTTCAGCTTCCT 5924
QY 1673 CTGCTGCACCAAGTCTTCACTTTGGGCAAGAGGAACTTTTGGAGAGGCGT 1732
DB 5923 CTGCTGCACCAAGTCTTCACTTTGGGCAAGAGGAACTTTTGGAGAGGCGT 5864
QY 1733 CTGCTGTGTACCCAGCTGGAATGCAAGTGGCGGATCTCACTCAACCGCACTTCAC 1792
DB 5863 CTGCTGTGTACCCAGCTGGAATGCAAGTGGCGGATCTCACTCAACCGCACTTCAC 5804
QY 1793 CTCTGGGTTCAAGTATTTTCTGCTCAAGCCCTCCCAAGTGGGATTAACGAGAC 1852
DB 5803 CTCTGGGTTCAAGTATTTTCTGCTCAAGCCCTCCCAAGTGGGATTAACGAGAC 5744
QY 1853 CCAACATGCCAGCTAATTTTGTATTTTCAAGTGAAGCGGATTTTCAACAGTTGGCCA 1912
DB 5743 CCAACATGCCAGCTAATTTTGTATTTTCAAGTGAAGCGGATTTTCAACAGTTGGCCA 5684
QY 1913 GGTGTGTCTGAATCTCTGACCGCAAGTATCAACCGCTCCGCTCCCAAGTGTGG 1972
DB 5683 GGTGTGTCTGAATCTCTGACCGCAAGTATCAACCGCTCCGCTCCCAAGTGTGG 5624
QY 1973 GATTACAGGGGTGAGCACCAGTGCAGGCGCCCAAGGGGAAACTTTTGGAGAGAGAGA 2032
DB 5623 GATTACAGGGGTGAGCACCAGTGCAGGCGCCCAAGGGGAAACTTTTGGAGAGAGAGA 5564
QY 2033 GGGGCTCAATCTCTCTGATTTCCCAATGACATTTGCTTCTCCCATTTAGC 2092
DB 5563 GGGGCTCAATCTCTCTGATTTCCCAATGACATTTGCTTCTCTCCCATTTAGC 5504

QY 2093 CAGGAATCTATTGTGTTTCTTCTGCAATTTACTATGATGTGATGCGGCTACCA 2152
DB 5503 CAGGAATCTATTGTGTTTCTTCTGCAATTTACTATGATGTGATGCGGCTACCA 5444
QY 2153 CCACCCCCCAGAGGGGGGTGAGAGGGGTGCAAGCCCTGCTCCACTTTTCTA 2212
DB 5443 CCACCCCCCAGAGGGGGGTGAGAGGGGTGCAAGCCCTGCTCCACTTTTCTA 5384
QY 2213 CCTTGAAGTGTATTTAGATAAATCACTTCTGTTGTTCAAGTTTCA 2260
DB 5383 CCTTGAAGTGTATTTAGATAAATCACTTCTGTTGTTCAAGTTTCA 5336
RESULT 13
ABAI9477
ID ABAI9477 standard; DNA; 32174 BP.
XX ABAI9477;
AC
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 11808.
XX
KW Human; nootropic; neuroprotective; cyrostatic; dermatological; virocidic;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;
KW antihemetic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antileptic; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotoxic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN MO200159063-A2.
XX
PD 16-AUG-2001.
XX
XX 17-JAN-2001; 2001MO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.

PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234927.
PR 25-SEP-2000; 2000US-0234928.
PR 25-SEP-2000; 2000US-0234929.
PR 26-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236328.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246529.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246529.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246530.
PR 08-NOV-2000; 2000US-0246510.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
XX Disclosure; SEQ ID NO 11808; 1701bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABA11004-ABA21513) and proteins
CC (ABA14678-ABA18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 32174 BP; 8747 A; 7064 C; 6837 G; 8001 T; 1525 other;

Query Match 31.1%; Score 708; DB 22; Length 32174;
Best Local Similarity 100.0%; Pred. No. 2.8e-300;
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1553 GGTGGCTGTGGGAGCTGTGCAGAAATCTACTGCTCTTTCACAGCACTCTTGG 1612
DB 26132 GGTGGCTGTGGGAGCTGTGCAGAAATCTACTGCTCTTTCACAGCACTCTTGG 26191
QY 1613 CCCAGAGCAGAGAAATGAGAAAGCCAGGAGGTGAGATGATGCTTCCAGCTGTGCTT 1672

Db 26192 CCCAGAGCAGAAATGAAAGCAGGAGGTGAAGATGATGCTTCCAGCTGTGCT 26251
Qy 1673 CTGCTGCACGACCAAGTCTTCATTTGGGCGCAAGGGGAACTTTTGGAGAGGCGT 1732
Db 26252 CTGCTGCACGACCAAGTCTTCATTTGGGCGCAAGGGGAACTTTTGGAGAGGCGT 26311
Qy 1733 CTGCTTGTTCACCGACGTGGAATGAGTGGGGATCTCAGTTCACCGCACTTCAC 1792
Db 26312 CTGCTTGTTCACCGACGTGGAATGAGTGGGGATCTCAGTTCACCGCACTTCAC 26371
Qy 1793 CTGCTGGGTCAAGTATTTCCGCTCAGCCTCCCAAGTAGTGGGAATACAGGACG 1852
Db 26372 CTGCTGGGTCAAGTATTTCCGCTCAGCCTCCCAAGTAGTGGGAATACAGGACG 26431
Qy 1853 CCAACATGCCAGCTAATTTTGTATTTTCAGTAAGAAACGGGATTTCCACACGTTGCGCA 1912
Db 26432 CCAACATGCCAGCTAATTTTGTATTTTCAGTAAGAAACGGGATTTCCACACGTTGCGCA 26491
Qy 1913 GGTGTGTTGAACTCTGACCGCAATGATCCACCGCTCCGCTCCCAAGTGTGG 1972
Db 26492 GGTGTGTTGAACTCTGACCGCAATGATCCACCGCTCCGCTCCCAAGTGTGG 26551
Qy 1973 GATTACAGGGGTGAGCAGCGTGCAGGCGCCCAAGGGGAACTTTGGAGAGGACGA 2032
Db 26552 GATTACAGGGGTGAGCAGCGTGCAGGCGCCCAAGGGGAACTTTGGAGAGGACGA 26611
Qy 2033 GGGGCTCACAATCTCCCTGTGATCCCAATGCAATGCTTCTCCCATCTAGC 2092
Db 26612 GGGGCTCACAATCTCCCTGTGATCCCAATGCAATGCTTCTCCCATCTAGC 26671
Qy 2093 CAGGAATCTATTGTTTCTTCTGCAATTACATGATGATGATGCGGCTACCA 2152
Db 26672 CAGGAATCTATTGTTTCTTCTGCAATTACATGATGATGATGCGGCTACCA 26731
Qy 2153 CCAACCCCCCATGGGGGGGAGAGGGGTCAGAGCCCTGCTCCACTTTTCTA 2212
Db 26732 CCAACCCCCCATGGGGGGGAGAGGGGTCAGAGCCCTGCTCCACTTTTCTA 26791
Qy 2213 CCTTGAATGATTAATAAATCACTTCTGTTTGTTCAGTTTCA 2260
Db 26792 CCTTGAATGATTAATAAATCACTTCTGTTTGTTCAGTTTCA 26839

RESULT 14
ABA20359 standard; DNA; 32174 BP.
XX ABA20359;
AC ABA20359;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 12690.
XX
KW Human, noctropic; neuroprotective; cytosolic; dermatological; virucide;
KW immunosuppressive; anti-inflamatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antidiabetic; antianemic; antiallergic; cancer;
KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
XX Homo sapiens.
OS
XX
PN WO200159063-A2.
XX
XX 16-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US01334.
PF
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.

RESULT 15
ABA21505/c
ID ABA21505 standard; DNA; 32174 BP.
XX
AC ABA21505;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 13836.
XX
KW Human; noctropic; neuroprotective; cytostatic; dermatological; virocidic;
KW immunosuppressive; anti-infectious; anti-HIV; antibacterial; cancer;
KW antiparkinsonian; antidiabetic; antianemic; antitubercular; cancer;
KW antihemorrhagic; hepatocarcinoma; neuroprotective; anti-infectious;
KW antiallergic; antidiabetic; antitubercular; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotoxic; gene therapy; vaccine; de.
XX
OS Homo sapiens.
XX
PN MO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0198774.
PR 17-MAR-2000; 2000US-0198076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227187.
PR 23-AUG-2000; 2000US-0227189.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR

PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 11-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.

17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
XX Disclosure; SEQ ID NO 13836; 1701bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABAI1004-ABAI21534) and proteins
CC (ABAI4678-ABAI8001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 32174 BP; 8001 A; 6837 C; 7064 G; 8747 T; 1525 other;

Query Match 31.1%; Score 708; DB 22; Length 32174;
Best Local Similarity 100.0%; Pred. No. 2.8e-300;
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1553 GGTGGCTGTGGGGA CTGTGAGAACTACTCGTCTCCCTTTTACAGCACTCCTTG 1612
DB 6043 GGTGGCTGTGGGGA CTGTGAGAACTACTCGTCTCCCTTTTACAGCACTCCTTG 5984
QY 1613 CCCAGAGCAGAGATGAGAAAGCCAGGAGGTGAGATGATGCTTCCAGCTTGCT 1672
DB 5983 CCCAGAGCAGAGATGAGAAAGCCAGGAGGTGAGATGATGCTTCCAGCTTGCT 5924
QY 1673 CTGCTGCCAGCCCAAGTCTTCTTTGGGGCAAAAGGGGAAACTTTTGTGAGAGGCGT 1732
DB 5923 CTGCTGCCAGCCCAAGTCTTCTTTGGGGCAAAAGGGGAAACTTTTGTGAGAGGCGT 5864
QY 1733 CTGCTTTTGCACCCAGCGTGAATGAGTGGGGGATCTCAGCTCAGCGCAACTCCAC 1792
DB 5863 CTGCTTTTGCACCCAGCGTGAATGAGTGGGGGATCTCAGCTCAGCGCAACTCCAC 5804
QY 1793 CTCCTGGGTTCAGTATTTTCTGCTCAGCTCCCAAGTAGTGGGATACAGGACG 1852
DB 5804 CTCCTGGGTTCAGTATTTTCTGCTCAGCTCCCAAGTAGTGGGATACAGGACG 1852

DB 5803 CTCCTGGGTTCAGTATTTTCTGCTCAGCTCCCAAGTAGTGGGATACAGGACG 5744
QY 1853 CCACCATGCCCGCTAAATTTTGTATTTTCTAGTAAACGGGATTTTCCAGAGTGGCCA 1912
DB 5743 CCACCATGCCCGCTAAATTTTGTATTTTCTAGTAAACGGGATTTTCCAGAGTGGCCA 5684
QY 1913 GGCTGATCGAACTCTGACCGCAAGTATCACCCTCCGCTCCGCTCCCAAGTCTGG 1972
DB 5683 GGCTGATCGAACTCTGACCGCAAGTATCACCCTCCGCTCCGCTCCCAAGTCTGG 5624
QY 1973 GATTACAGCGGTGACCAACCGTCCCGGCCCAAGGGGAACTCTTGTGGAGAGCAGA 2032
DB 5623 GATTACAGCGGTGACCAACCGTCCCGGCCCAAGGGGAACTCTTGTGGAGAGCAGA 5564
QY 2033 GGGGCTCACATCTCCCTCTGATTTTCCCAATGATGCTTATCTCCCATCTAGC 2092
DB 5563 GGGGCTCACATCTCCCTCTGATTTTCCCAATGATGCTTATCTCCCATCTAGC 5504
QY 2093 CAGGAATCTATTTGTTTCTTCTGCAATTTACTATGATGATGATGCGGCTACCA 2152
DB 5503 CAGGAATCTATTTGTTTCTTCTGCAATTTACTATGATGATGATGCGGCTACCA 5444
QY 2153 CCACCCCGCCCATGCGGGGGGTGAGAGGGGTGCAAGGCCCTGCTCCACTTTTCTA 2212
DB 5443 CCACCCCGCCCATGCGGGGGGTGAGAGGGGTGCAAGGCCCTGCTCCACTTTTCTA 5384
QY 2213 CCTTGAATCTGATTAATAATATCTTCTTGTGTTGTTTCA 2260
DB 5383 CCTTGAATCTGATTAATAATATCTTCTTGTGTTGTTTCA 5336

Search completed: November 22, 2003, 10:43:39
Job time : 603 secs

THIS PAGE BLANK (USPTO)

Db 61 GGGCTCCCGAGATTAGATGGCGTCTCAGCGGAGGGGAGCAGAGGGAAGCTGTGTGGCC 120
Qy 121 TGGCGGGGGGTTCTGCACTCGGGTTTCCAGAGCTGAGCCTTAAACAAGTTGGCGAGCTGCC 180
Db 121 TGGCGGGGGGTTCTGCACTCGGGTTTCCAGAGCTGAGCCTTAAACAAGTTGGCGAGCTGCC 180
Qy 181 TGGGCGGCTGAGAAACAGGGGCTGGGGTATCATCTCCATCTTCCGGGTTACCCCTTGG 240
Db 181 TGGGCGGCTGAGAAACAGGGGCTGGGGTATCATCTCCATCTTCCGGGTTACCCCTTGG 240
Qy 241 CTATTGTTTATCGGCATATTACTTTTCTACAGAGAGACTCTCATCCACTCTTCCATA 300
Db 241 CTATTGTTTATCGGCATATTACTTTTCTACAGAGAGACTCTCATCCACTCTTCCATA 300
Qy 301 CCTTTACAGGCTCTCATTTGCTTATTATTAACTTTGAAAACAAGCTCTACCTCCCTGC 360
Db 301 CCTTTACAGGCTCTCATTTGCTTATTATTAACTTTGAAAACAAGCTCTACCTCCCTGC 360
Qy 361 TGTGTAATGTGCTCAGTTCCTCATCTGATCCGTAATGAGGCGGACACATCATGCGGTCG 420
Db 361 TGTGTAATGTGCTCAGTTCCTCATCTGATCCGTAATGAGGCGGACACATCATGCGGTCG 420
Qy 421 TCACATCCTTTTCTTCCAGATGAGCTTCTGCTGATATCTATTACATGCGACG 480
Db 421 TCACATCCTTTTCTTCCAGATGAGCTTCTGCTGATATCTATTACATGCGACG 480
Qy 481 GCAACTACGATATCAAGTGAACAATGCAATGCTGTCTGACTTTGAACTGATGTT 540
Db 481 GCAACTACGATATCAAGTGAACAATGCAATGCTGTCTGACTTTGAACTGATGTT 540
Qy 541 TGGCTGTGATCTATTGACGAGGAGGAAATGAGATTCCTTCTCTGACCAACA 600
Db 541 TGGCTGTGATCTATTGACGAGGAGGAAATGAGATTCCTTCTCTGACCAACA 600
Qy 601 AATATGCAATACGTGTGTCTCTCCCTGCTGAGAGTTGCTGTTCTCTACTTCTATG 660
Db 601 AATATGCAATACGTGTGTCTCTCCCTGCTGAGAGTTGCTGTTCTCTACTTCTATG 660
Qy 661 GGGGCTTTCTGTGAGGGGCCCCAGTTCTCAATGAATCACTACATGAAAGCTGTACAGGAG 720
Db 661 GGGGCTTTCTGTGAGGGGCCCCAGTTCTCAATGAATCACTACATGAAAGCTGTACAGGAG 720
Qy 721 AGCTGATGATACCAAGGAAATACCAACAGCATCTCTGCTCTCAAGCGCTGA 780
Db 721 AGCTGATGATACCAAGGAAATACCAACAGCATCTCTGCTCTCAAGCGCTGA 780
Qy 781 GTCTGGGCTTTTCTAATGAGGCTACACACTGTCAAGCCCCCATCAAGAAAGCT 840
Db 781 GTCTGGGCTTTTCTAATGAGGCTACACACTGTCAAGCCCCCATCAAGAAAGCT 840
Qy 841 ATCTCTCACTGAAGATATGACCAACCCCTTCTGCTTCCGCTGATGATGATGCTGA 900
Db 841 ATCTCTCACTGAAGATATGACCAACCCCTTCTGCTTCCGCTGATGATGATGCTGA 900
Qy 901 TCTGGGGCAAGTTTGTGCTGATCAAAATATGCACTGTTGGCTGTGACAGAAAGGATAT 960
Db 901 TCTGGGGCAAGTTTGTGCTGATCAAAATATGCACTGTTGGCTGTGACAGAAAGGATAT 960
Qy 961 GCATTTTGAAGGGCTGTGCTCAATGCTTTTGAAGAAAGGCAAGGCAAGTGGGATG 1020
Db 961 GCATTTTGAAGGGCTGTGCTCAATGCTTTTGAAGAAAGGCAAGGCAAGTGGGATG 1020
Qy 1021 CCTGTGCAACATGAAAGGTGTGCTTTTGAACAAACCCCGCTTCACTGAGCACTATG 1080
Db 1021 CCTGTGCAACATGAAAGGTGTGCTTTTGAACAAACCCCGCTTCACTGAGCACTATG 1080
Qy 1081 CCTCATTCACATCAACACAGCGCTGGGTGGCGCTACATCTTCAACGACTCAAGT 1140
Db 1081 CCTCATTCACATCAACACAGCGCTGGGTGGCGCTACATCTTCAACGACTCAAGT 1140
Qy 1141 TCTTGGAAATTAAGAACTCTCTCAGGGGTCTCTGTTGCTATTCCTGGCCCTCTGGACG 1200
Db 1141 TCTTGGAAATTAAGAACTCTCTCAGGGGTCTCTGTTGCTATTCCTGGCCCTCTGGACG 1200

Qy 1201 GCCTGACTCAGAGTACCTGTGTCTTCCAGATGGAATTCCTCATTTGTTATTTGGAAA 1260
Db 1201 GCCTGACTCAGAGTACCTGTGTCTTCCAGATGGAATTCCTCATTTGTTATTTGGAAA 1260
Qy 1261 GACAGGCTGCAGGCTCATTTCAAGAGAGGCCACCCCTGAGCAAGCTGGCGGCACTACTG 1320
Db 1261 GACAGGCTGCAGGCTCATTTCAAGAGAGGCCACCCCTGAGCAAGCTGGCGGCACTACTG 1320
Qy 1321 TCCCTCAGGCTTCTAATTTTGTGCAACAGACATCCACTGACTTTTCAATGGATTACT 1380
Db 1321 TCCCTCAGGCTTCTAATTTTGTGCAACAGACATCCACTGACTTTTCAATGGATTACT 1380
Qy 1381 CCATGACTGCTTCTGCTTCTTACGTGGGACAAATGCTTAAAGGTATTAATCCATCT 1440
Db 1381 CCATGACTGCTTCTGCTTCTTACGTGGGACAAATGCTTAAAGGTATTAATCCATCT 1440
Qy 1441 ATTTCCTGAGCACATCTTCTCCTGAGGCTACTATTCATATTCCTTATTAATCACAAG 1500
Db 1441 ATTTCCTGAGCACATCTTCTCCTGAGGCTACTATTCATATTCCTTATTAATCACAAG 1500
Qy 1501 CAATGTGCAAGAGAAAGAAATTAAAGAAATGGAATTAATTCATTTCCCTGTGGCT 1560
Db 1501 CAATGTGCAAGAGAAAGAAATTAAAGAAATGGAATTAATTCATTTCCCTGTGGCT 1560
Qy 1561 GTGCGGACTGTGTGCAAGAACTACTGCTTCTCTTTTCAAGACATCTCTTGGCCAGAG 1620
Db 1561 GTGCGGACTGTGTGCAAGAACTACTGCTTCTCTTTTCAAGACATCTCTTGGCCAGAG 1620
Qy 1621 CAGAGAAATGAAAAGCCAGAGAGGTGGAAGATCGATCTTCCAGCTGTGCTGTGCTGC 1680
Db 1621 CAGAGAAATGAAAAGCCAGAGAGGTGGAAGATCGATCTTCCAGCTGTGCTGTGCTGC 1680
Qy 1681 AGCAAGTCTTCAATTTGGGGCCAAAGGGGAAAATTTTTTGTGAGAAAGCGTCTTCT 1740
Db 1681 AGCAAGTCTTCAATTTGGGGCCAAAGGGGAAAATTTTTTGTGAGAAAGCGTCTTCT 1740
Qy 1741 GTACCCCAAGCTGGAATGCAAGTGGGGATCTCAGCTCACCGCACTCTCTCTGG 1800
Db 1741 GTACCCCAAGCTGGAATGCAAGTGGGGATCTCAGCTCACCGCACTCTCTCTGG 1800
Qy 1801 TTCAAGTATTTTCTGCTCAGGCTCCCAAGAGTGGGAAATACAGGACGCAACATG 1860
Db 1801 TTCAAGTATTTTCTGCTCAGGCTCCCAAGAGTGGGAAATACAGGACGCAACATG 1860
Qy 1861 CCCAGCTAATTTTGTATTTTCAAGTAAACGGGATTTTCAACAGCTGTGCGCAGGCTGTC 1920
Db 1861 CCCAGCTAATTTTGTATTTTCAAGTAAACGGGATTTTCAACAGCTGTGCGCAGGCTGTC 1920
Qy 1921 TCGAATCTCTGACCGCAAGTATCCACCGGCTTCCGCTTCCCAAGTGTGGGATTAACG 1980
Db 1921 TCGAATCTCTGACCGCAAGTATCCACCGGCTTCCGCTTCCCAAGTGTGGGATTAACG 1980
Qy 1981 GCGTAGCACCAGTGTGCGGAGCCCAAGAGGGAATCTTGTGAGAGGAGAGAGGCTCA 2040
Db 1981 GCGTAGCACCAGTGTGCGGAGCCCAAGAGGGAATCTTGTGAGAGGAGAGAGGCTCA 2040
Qy 2041 CATCTCCCTGTGATTCCTCCATGACATTTGCTTCTCCCATAGCCAGGAATC 2100
Db 2041 CATCTCCCTGTGATTCCTCCATGACATTTGCTTCTCCCATAGCCAGGAATC 2100
Qy 2101 TATTTGTTTTTCTTCTGCAATTTATGATTTGTGTAATGTGCGGCTACACACACCCC 2160
Db 2101 TATTTGTTTTTCTTCTGCAATTTATGATTTGTGTAATGTGCGGCTACACACACCCC 2160
Qy 2161 CCCATGGGGGGGTGAGAGGGGTGCAAGGCGCTGCTCCACTTTTCTACCTTGGAA 2220
Db 2161 CCCATGGGGGGGTGAGAGGGGTGCAAGGCGCTGCTCCACTTTTCTACCTTGGAA 2220
Qy 2221 CTGTATTAATAAATCACTTCTGTTGTTCAGTTTTCAAAAA 2273
Db 2221 CTGTATTAATAAATCACTTCTGTTGTTCAGTTTTCAAAAA 2273

RESULT 2
US-09-880-107-3415
; Sequence 3415, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT FILING DATE: 2001-06-14
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3415
; LENGTH: 1842
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U72515
US-09-880-107-3415

Query Match 81.0%; Score 1842; DB 10; Length 1842;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 TACCCATCCACCTCTTCATACCTTTACAGGCTCTCAATGCTTATTTTACCTTGA 338
DB 1 TACCTATCCACCTCTTCATACCTTTACAGGCTCTCAATGCTTATTTTACCTTGA 60
QY 339 AACGACTTACACCTCTCTGCTGATATGCTTCACTTCTCATCTTCTGATTAAG 398
DB 61 AACGACTTACACCTCTCTGCTGATATGCTTCACTTCTCATCTTCTGATTAAG 120
QY 339 GGGCCACCATACCTGCGCTCTCTACCTTTTGGCTTCAGATGAGCTTCTTGCT 458
DB 121 GGGCCACCATACCTGCGCTCTCTACCTTTTGGCTTCAGATGAGCTTCTTGCT 180
QY 459 GGATCTATTACCTGCGCAGGCACTACATCAAGTGAATGCAATGCTTGT 518
DB 181 GGATCTATTACCTGCGCAGGCACTACATCAAGTGAATGCAATGCTTGT 240
QY 519 CTGACTTTGAAGCTATGTTGGCTGTGACTTCTTGAAGGAGGAAAGATCAGAT 578
DB 241 CTGACTTTGAAGCTATGTTGGCTGTGACTTCTTGAAGGAGGAAAGATCAGAT 300
QY 579 TCCCTGCTCTGAGCAACAGAAATATGCAATGCTGCTTCTTCCCTGCTGAAGT 638
DB 301 TCCCTGCTCTGAGCAACAGAAATATGCAATGCTGCTTCTTCCCTGCTGAAGT 360
QY 639 GCTGCTTCTCTACTTCTATGAGGCTTCTTGTGAGGCGCCCAATTCATGATC 698
DB 361 GCTGCTTCTCTACTTCTATGAGGCTTCTTGTGAGGCGCCCAATTCATGATC 420
QY 699 TACATGAAGCTGTCAGGAGAGCTGATTTGACATACAGAAATACCAAGCAGATC 758
DB 421 TACATGAAGCTGTCAGGAGAGCTGATTTGACATACAGAAATACCAAGCAGATC 480
QY 759 ATTCTGCTCTCAAGGCTGAGTCTGGGCTTTTCTTACCTTATGAGGCTTCACTGCTC 818
DB 481 ATTCTGCTCTCAAGGCTGAGTCTGGGCTTTTCTTACCTTATGAGGCTTCACTGCTC 540
QY 819 AGCCCAATCATCAGAGAGCTATCTCTCACTGAAGATATGACAAACCCCTTGG 878
DB 541 AGCCCAATCATCAGAGAGCTATCTCTCACTGAAGATATGACAAACCCCTTGG 600
QY 879 TTCCTGCTGATGATGCTGATCTGGGCAAGTTTGTGCTGATCAAAATATGCTGCTGT 938

DB 601 TTCGCTCATGTATCATGCTGATCTGGGCAAGTTTGTGCTGATCAAAATATGCTGCTGT 660
QY 939 TGGCTGTCACAGAGAGATATGATTTTGAAGGCGCTTGGCTTCAATGCTTTGAAGA 998
DB 661 TGGCTGTCACAGAGAGATATGATTTTGAAGGCGCTTGGCTTCAATGCTTTGAAGA 720
QY 999 AAGGCAAGGCAAGAGTGGATGCTGTCACATGAGAGTGTGGCTTTTGAACAAC 1058
DB 721 AAGGCAAGGCAAGAGTGGATGCTGTCACATGAGAGTGTGGCTTTTGAACAAC 780
QY 1059 CCCCCTTCACTGACCATTTGCTCATTAACATCAACCAACGCTGAGGCGCCG 1118
DB 781 CCCCCTTCACTGACCATTTGCTCATTAACATCAACCAACGCTGAGGCGCCG 840
QY 1119 TACATCTTCAACAGACTAAGTCTTGTGAATAAAGAACTCTCTGAGGCTCTGCTG 1178
DB 841 TACATCTTCAACAGACTAAGTCTTGTGAATAAAGAACTCTCTGAGGCTCTGCTG 900
QY 1179 CTATTCCTGAGGCTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1238
DB 901 CTATTCCTGAGGCTCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 960
QY 1239 TTCTCATTTTATTTGGAAGAAGAGGCTGAGGCTCATTCAGAGGCGCCGCTG 1298
DB 961 TTCTCATTTTATTTGGAAGAAGAGGCTGAGGCTCATTCAGAGGCGCCGCTG 1020
QY 1299 AGCAAGCTGCGGCGCATTAAGTCTCTGAGGCTCTTCTATTTGCTGCAACAGCATC 1358
DB 1021 AGCAAGCTGCGGCGCATTAAGTCTCTGAGGCTCTTCTATTTGCTGCAACAGCATC 1080
QY 1359 CACTGAGCTTCAAGGCTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1418
DB 1081 CACTGAGCTTCAAGGCTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1140
QY 1419 CTTAAGTGTATTAATCATCTATTTCTTGGCCACATCTTCTTCTGAGCTACTATTC 1478
DB 1141 CTTAAGTGTATTAATCATCTATTTCTTGGCCACATCTTCTTCTGAGCTACTATTC 1200
QY 1479 ATATTGCTTATTTACAAAGCAATGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1538
DB 1201 ATATTGCTTATTTACAAAGCAATGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1539 TAATCATTTTCCCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1598
DB 1261 TAATCATTTTCCCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1320
QY 1599 ACAGCACTCTTTTGGCCAG 1658
DB 1321 ACAGCACTCTTTTGGCCAG 1380
QY 1659 TTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1718
DB 1381 TTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
QY 1719 TTTGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1778
DB 1441 TTTGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
QY 1779 ACCGCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1838
DB 1501 ACCGCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
QY 1839 GGAATACAGGCAAGGCAAGTCCAGTAAATTTTGTATTTTCAATGAAAGGGAATTT 1898
DB 1561 GGAATACAGGCAAGGCAAGTCCAGTAAATTTTGTATTTTCAATGAAAGGGAATTT 1620
QY 1899 CACCAAGTGGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1958
DB 1621 CACCAAGTGGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
QY 1959 TCCCAAGTGGCAAGTTCAGGCTGAGGCAAGTTCAGGCTGAGGCAAGTTCAGGCTGAG 2018

Db 1681 TCCCAAGTCTGGATTAACAGGCGTGGACCGCCGCCCAAGGGAACCTTT 1740
Qy 2019 GTGGAGAGACAGAGGCGCTCACATCTCCCTTGATTTCCCATGACATTGCTTATC 2078
Db 1741 GTGGAGAGACAGAGGCGCTCACATCTCCCTTGATTTCCCATGACATTGCTTATC 1800
Qy 2079 TCTCCCATCTAGCAGGAATCTATTGTGTTTCTTCTTCCG 2120
Db 1801 TCTCCCATCTAGCAGGAATCTATTGTGTTTCTTCTTCCG 1842

RESULT 3
US-09-954-531-1368
Sequence 1368, Application US/09954531
Patent No. US20020165180A1
GENERAL INFORMATION:
APPLICANT: Weaver, Zoe
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
TITLE OF INVENTION: Gene Sets
FILE REFERENCE: 689290-77
CURRENT APPLICATION NUMBER: US/09/954,531
CURRENT FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: US/60/233,133
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,009
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,034
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,509
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: US/60/234,567
PRIOR FILING DATE: 2000-09-22
NUMBER OF SEQ ID NOS: 1392
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1368
LENGTH: 1842
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-531-1368

Query Match 81.0%; Score 1842; DB 10; Length 1842;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 279 TACCTCATCACTCTTCCATACCTTTACAGGCGCTTCAATTGCTTATTTAACTTGA 338
Db 1 TACCTCATCACTCTTCCATACCTTTACAGGCGCTTCAATTGCTTATTTAACTTGA 60
Qy 339 AACCAAGCTACCACTCCGCTGTGTATGCTTCAATGCTTCACTTCCACTTGA 398
Db 61 AACCAAGCTACCACTCCGCTGTGTATGCTTCAATGCTTCACTTCCACTTGA 120
Qy 399 GGGCGACATCACTGCGCTCTCTCACTACCTTTTGTCCAGATGGCTTACCTTGGCT 458
Db 121 GGGCGACATCACTGCGCTCTCTCACTACCTTTTGTCCAGATGGCTTACCTTGGCT 180
Qy 459 GGATCACTTACATCTGCGACCGGCACTAGATATCAAGTGGAGAAATGCCATTTGTT 518
Db 181 GGATCACTTACATCTGCGACCGGCACTAGATATCAAGTGGAGAAATGCCATTTGTT 240
Qy 519 CTGACTTTGAAGCTGATTTGTTGGCTGTGACTACTTTCAGGAGGAAAGATCAGAT 578
Db 241 CTGACTTTGAAGCTGATTTGTTGGCTGTGACTACTTTCAGGAGGAAAGATCAGAT 300
Qy 579 TCTTGTCTCTGAGCAAGAAATATGCCATAGTGTGTTCCCTTCCGCTGGAGATT 638
Db 301 TCTTGTCTCTGAGCAAGAAATATGCCATAGTGTGTTCCCTTCCGCTGGAGATT 360
Qy 639 GCTGTGTTCTCTATCTTATGAGGCGCTTCTTGTGAGGCGCCAGTTCTCAATTAATCAC 698
Db 361 GCTGTGTTCTCTATCTTATGAGGCGCTTCTTGTGAGGCGCCAGTTCTCAATTAATCAC 420
Qy 699 TACATGAAGCTGTGAGGAGAGCTGATTTGACATACAGGAAGATACCAACAGCATC 758

Db 421 TACATGAAGCTGTGAGGAGAGCTGATTTGACATACAGGAAGATACCAACAGCATC 480
Qy 759 ATTCTGCTCTGAGGCGCTGAGTGTGGCGCTTTTCTACCTAGTGGGCTACACATGCTC 818
Db 481 ATTCTGCTCTGAGGCGCTGAGTGTGGCGCTTTTCTACCTAGTGGGCTACACATGCTC 540
Qy 819 AGCCCCCATACAGAAAGATCTATCTCTCATGGAAGACTATAGAACCAACCCCTTCTGG 878
Db 541 AGCCCCCATACAGAAAGATCTATCTCTCATGGAAGACTATAGAACCAACCCCTTCTGG 600
Qy 879 TTCCGCTGATGATGATGCTGATCTGAGGCAAGTTTGTCTGATCAATATGTCACCTGT 938
Db 601 TTCCGCTGATGATGATGCTGATCTGAGGCAAGTTTGTCTGATCAATATGTCACCTGT 660
Qy 939 TGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 998
Db 661 TGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
Qy 999 AAGGCGAAGCAAGTGGGATGCTGTGCAACATGAAGTGTGCTTTTGAACAAAC 1058
Db 721 AAGGCGAAGCAAGTGGGATGCTGTGCAACATGAAGTGTGCTTTTGAACAAAC 780
Qy 1059 CCCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1118
Db 781 CCCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
Qy 1119 TACATCTTCAAGACTCAAGTCTCTTGAAGATTAAGAACTCTGAGGCTCTCTGCTT 1178
Db 841 TACATCTTCAAGACTCAAGTCTCTTGAAGATTAAGAACTCTGAGGCTCTCTGCTT 900
Qy 1179 CTATTCCTGCGCTCTGAGACGGCTGCACTAGATTAACCTGCTCTTCAAGATGGA 1238
Db 901 CTATTCCTGCGCTCTGAGACGGCTGCACTAGATTAACCTGCTCTTCAAGATGGA 960
Qy 1239 TTCTCATTTGATTTGATGAAAGACAGGCTGCCAGCTCATTCAGAGCCCACTTGC 1298
Db 961 TTCTCATTTGATTTGATGAAAGACAGGCTGCCAGCTCATTCAGAGCCCACTTGC 1020
Qy 1299 AGCAAGCTGGCGCATTAATCTCTTCAAGCTTCTTCAATTTGATGCAAGACCATC 1358
Db 1021 AGCAAGCTGGCGCATTAATCTCTTCAAGCTTCTTCAATTTGATGCAAGACCATC 1080
Qy 1359 CACTGCTCTTCAATGATTAATCAATGATGATGATGATGATGATGATGATGATGATG 1418
Db 1081 CACTGCTCTTCAATGATTAATCAATGATGATGATGATGATGATGATGATGATGATG 1140
Qy 1419 CTTAAGGTGATTAATCAATCTATTTCTTGGCCATCATCTTCTCTGAGCTTACTATTC 1478
Db 1141 CTTAAGGTGATTAATCAATCTATTTCTTGGCCATCATCTTCTCTGAGCTTACTATTC 1200
Qy 1479 ATATGCTTTATTAATCAAGAAAGATGATGATGATGATGATGATGATGATGATGATG 1538
Db 1201 ATATGCTTTATTAATCAAGAAAGATGATGATGATGATGATGATGATGATGATGATG 1260
Qy 1539 TAATCAATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1598
Db 1261 TAATCAATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1320
Qy 1599 ACAGACTCTTTTCCCGAGAGCAAGATGAAAGCCAGGAGGTGAAAGATGATGATG 1658
Db 1321 ACAGACTCTTTTCCCGAGAGCAAGATGAAAGCCAGGAGGTGAAAGATGATGATG 1380
Qy 1659 TTCCAGCTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1718
Db 1381 TTCCAGCTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
Qy 1719 TTGGAAGAGGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1778
Db 1441 TTGGAAGAGGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
Qy 1779 ACCGAACTCTCACTCTGAGGTTCAAGTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1838

Db	1501	ACCGAACTCACCTCTGGGTTCAAGTATTTTCGCTCAAGCTCCCAAGTAGCTG	1566
Qy	1839	GGATATACAGGCACCGCACCATGCGCAGTAAATTTTTGTATTTTCAGTAGAAACGGGATTT	18999
Db	1561	GGAAATACAGGACCGCACCATGCGCAGGATTAATTTTTGTATTTTCAGTAGAAACGGGATTT	1620
Qy	1899	CACCACTGTGGCCAGGCTGTGTCGAACTCTGACCGCAAGTATCCACCGCTCGGCC	1958
Db	1621	CACCACTGTGGCCAGGCTGTGTCGAACTCTGACCGCAAGTATCCACCGCTCGGCC	1680
Qy	1593	TCCCAATGCTGCGATTAACAGGCGTAGCCACCGTGCCTGGCCCAAGGGGAACTTT	2018
Db	1661	TCCCAATGCTGCGATTAACAGGCGTAGCCACCGTGCCTGGCCCAAGGGGAACTTT	1740
Qy	2019	GTGGAGAGCAGAGGGGCTACATCTCCCTGTATTCGCCCATGACATTTGCTTATC	20787
Db	1741	GTGGAGAGCAGAGGGGCTACATCTCCCTGTATTCGCCCATGACATTTGCTTATC	18000
Qy	2079	TCTCCCATCTAGCCAGAAATCTATGTGTATTTTCTTTGCTG	2120
Db	1801	TCTCCCATCTAGCCAGAAATCTATGTGTATTTTCTTTGCTG	1842

```

RESULT 4
US-09-764-860-1135/c
; Sequence 1135, Application US/09764860
; Patent No US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; PRIORITY FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1135
; -LENGTH: 7461
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-860-1135

```

Query Match	Similarity	31.2%	Score 708.6	DB 9	Length 7461
Best Local Match	Similarity	99.4%	Pred. No. 2.8e-209		
Matches 711	Conservative	0	MisMatches 4	Indels 0	Gaps 0
OY	1546	TTTCCTGTGAGCTGTGCGGAGCTGCGAGAACTATCTGCTCCCTTTTCACAGAC	1609		
Db	721	TCTTCAGAGTGGCCGTGTGGGGAGCTGTGAGAACTATCTGTCTCCCTTTTCACAGAC	662		
OY	1606	TCCTTTGGCCCCAGACAGAGATGAAAAAGCCAGGGAGGTGAGATGATGCTTCCAGC	1665		
Db	661	TCCTTTGGCCCCAGACAGAGATGAAAAAGCCAGGGAGGTGAGATGATGCTTCCAGC	602		
OY	1666	TGTGCTCTGCTGCGCAGCCAAAGTCTTCACTTTTGGGGCCAAAGGGGAAACTTTTTTTGGAG	1729		
Db	601	TGTGCTCTGTGCTGCCAGCCAACTCTCACTTTGGGGCCAAAGGGGAAACTTTTTTTGGAG	542		
OY	1726	AAGCGCTTGTGTTGTACCCACGCTGAGATGACAGTGGCGGGATCTCAGCTCACCGCA	1785		
Db	541	AAGCGCTTGTGTTGTACCCACGCTGAGATGACAGTGGCGGGATCTCAGCTCACCGCA	482		
OY	1786	CCTCACCCTCGGGGTTCAAGGATTTTTCCTGCGCCACGCTCCCAAGTAGCTGGGGAATAC	1845		
Db	481	CCTCACCCTCTGGGTTCAAGGATTTTTCCTGCGCCACGCTCCCAAGTAGCTGGGGAATAC	422		
OY	1846	AGGACGCGCACCATGCCAGCTAAATTTTGTATTTTCAGTAGAAAGGGATTTCCACAG	1905		
Db	421	AGGACGCGCACCATGCCAGCTAAATTTTGTATTTTCAGTAGAAAGGGATTTCCACAG	362		
OY	1906	TTGGCAGAGGTGTGTCGAATCTTCGACCGCAAGATTCACCGCCCTTCGGCTCCCAA	1965		
Db	361	TTGGCAGAGGTGTGTCGAATCTTCGACCGCAAGATTCACCGCCCTTCGGCTCCCAA	302		

QY	1986	GTGCTGGGATTAACAAGGGGTAGCCACCGCTGCCCCGGCCAAAAGGGGAAACTTTGTGGGAG	2025
Db	301	GTGCTGGGATTAACAGGCGTGAACCAACCGTCCGGCCAAAAGGGGAAACTTTGTGGGAG	242
QY	2026	GAGCAGAGGGGCTCACATCTCCCTCTGATTCGCCCATGACCATTCGCTTATCTCTCCC	2085
Db	241	GAGCAGAGGGGCTCACATCTCCCTCTGATTCGCCCATGACCATTCGCTTATCTCTCCC	182
QY	2086	ATCTAGCCAGGAATCTATGTGTTTCTTCTGTGCAATTACTATGATGTGATGAGCC	2145
Db	181	ATCTAGCCAGGAATCTATGTGTTTCTTCTGTGCAATTACTATGATGTGATGAGCC	122
QY	2146	GCTACCAACACCCCCCCCCATGSGGGGGGTGAGAGGGGTGCAAGGCCCTGCTGTCACT	2205
Db	121	GCTACCAACACCCCCCCCCATGSGGGGGGTGAGAGGGGTGCAAGGCCCTGCTGTCACT	62
QY	2206	TTTCTACCTTGGAACCTATTAAGTAATAACACTCTGTTTGTTGTCAGTTTTTCA	2260
Db	61	TTTCTACCTTGGAACCTATTAAGTAATAACACTCTGTTTGTTGTCAGTTTTTCA	7

```

RESULT 5
US-09-764-904-91/c
: Sequence 91, Application US/09764904
: Patent No. US20020173454A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PA122
: CURRENT APPLICATION NUMBER: US/09/764,904
: PRIORITY FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 137
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO. 91
: LENGTH: 7461
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-764-904-91

```

Query Match	Similarity	31.2%	Score 708.6;	DB 10;	Length 7461;
Beer Local	Similarity	99.4%	Pred. No.2.8e-209;		
Matches	711;	Conservative	0;	Mismatches	4;
				Indels	0;
				Gaps	0;
QY	1546	TTTCCTGTGAGCTGCTGTGCGGAGCTGTGCAAGAACTACTGCTCTCCCTTTTCACAGC	1605		
Db	721	TCTTSCAGAGTGGCTGTGCGGAGCTGTGCAAGAACTACTGCTCTCCCTTTTCACAGC	662		
QY	1606	TCCTTTGCCCCAGACAGAGATGAGAAAAGCCAGGAGGTGAGATGTGCTTCACG	1665		
Db	661	TCCTTTGCCCCAGACAGAGATGAGAAAAGCCAGGAGGTGAGATGTGCTTCACG	602		
QY	1666	TGTGCTCTGTGCTGACGAGCCAAAGTCTTCAATTGTGGGSCAAAGGGGAAACTTTTTTTGGAG	1726		
Db	601	TGTGCTCTGTGCTGACGAGCCAAAGTCTTCAATTGTGGGSCAAAGGGGAAACTTTTTTTGGAG	542		
QY	1726	AAGGCGTCTGTGTTGTTCACCCACGCTGAAATGCAAGTGAGTGCGGGATCTCAGCTCACGCCAA	1788		
Db	541	AAGGCGTCTGTGTTGTTCACCCACGCTGAAATGCAAGTGAGTGCGGGATCTCAGCTCACGCCAA	482		
QY	1786	CCTCCAGCTCGGGGTTCAAGGATTTTCTGCGCTCAGGCTCCCAAGTAGCTGGGAAATAC	1841		
Db	481	CCTCCAGCTCGGGGTTCAAGGATTTTCTGCGCTCAGGCTCCCAAGTAGCTGGGAAATAC	422		
QY	1846	AGGCAAGCCACCATGCGCCAGCTAATTTTGTATTTTCAGTAGAAACGGGATTTCCACG	1905		
Db	421	AGGCAAGCCACCATGCGCCAGCTAATTTTGTATTTTCAGTAGAAACGGGATTTCCACACG	362		
QY	1906	TTGSCCAGAGCTGTGTGCAATCTCTGACCCGCAAGATTCACCCGCGCTCTCCGCTCCCAA	1965		
Db	361	TTGSCCAGAGCTGTGTGCAATCTCTGACCCGCAAGATTCACCCGCGCTCTCCGCTCCCAA	302		


```

PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/236,369
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/224,519
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/220,964
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/241,809
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/249,299
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/236,327
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/241,785
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/244,617
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: 60/225,268
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/236,368
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/251,856
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/251,868
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/229,344
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/234,997
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: 60/229,343
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,345
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,287
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,513
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/231,413
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/229,509
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/236,367
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/237,039
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,038
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/236,370
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/236,802
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,037
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,040
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/240,960
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/239,935
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/239,937
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/241,787
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,474
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/246,532
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/249,216
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,210
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/226,681
PRIOR FILING DATE: 2000-08-22

```

```

PRIOR APPLICATION NUMBER: 60/225,759
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/225,213
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/227,182
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,214
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/235,836
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: 60/230,438
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/215,135
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: 60/225,266
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/249,218
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,208
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,213
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,212
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,207
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,245
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,244
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,217
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,211
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,215
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,264
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,214
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,297
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/232,400
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/231,242
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,081
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,080
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/231,414
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/231,244
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/233,064
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/233,063
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,397
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,399
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,401
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/241,808
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,826
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,786
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,221
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,475

```

```
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match      31.2%; Score 708.6; DB 14; Length 7461;
Best Local Similarity 99.4%; Pred. No. 2,8e-209;
Matches 711; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1546 TTTCCCTGTGGGCTGTGGGGGACTGTGTGAGAACTACTGCTCTCTTTTACAGCAC 1605
DB 721 TCTTGGAGGTGGCTGTGGGGGCTGTGTGAGAACTACTGCTCTCTTTTACAGCAC 662
QY 1606 TCTTTTGGCCCAAGAGAGAAATGAGAAAGCCAGGAGGAGGAGATGCTTCACAG 1665
DB 661 TCTTTTGGCCCAAGAGAGAAATGAGAAAGCCAGGAGGAGGAGATGCTTCACAG 602
QY 1666 TGTGCTCTGTGCTGCAAGCTCTTCAATTGGGGCCAAAGGGGAACTTTTTTGGAG 1725
DB 601 TGTGCTCTGTGCTGCAAGCTCTTCAATTGGGGCCAAAGGGGAACTTTTTTGGAG 542
QY 1726 AAGGCTCTGTGCTTGTGTACCCAGCTGGAATGCAATGCTGAGTCTCAGCTCACGCA 1785
DB 541 AAGGCTCTGTGCTTGTGTACCCAGCTGGAATGCAATGCTGAGTCTCAGCTCACGCA 482
QY 1786 CCTCCACTCTCTGGGTTCAAGTATTTTCTGCTCAGCTCCCAAGTAGCTGGGAATAC 1845
DB 481 CCTCCACTCTCTGGGTTCAAGTATTTTCTGCTCAGCTCCCAAGTAGCTGGGAATAC 422
QY 1846 AAGGAGCCAGCAATGCCCAGCTAATTTTGTATTTTCAAGTAAGCGGATTTTACACAG 1905
DB 421 AAGGAGCCAGCAATGCCCAGCTAATTTTGTATTTTCAAGTAAGCGGATTTTACACAG 362
QY 1906 TTGGCCAGGCTGTGCTGCAATCTCTGACCGCAAGTATCACCGGCTCTCGCTCCCAAA 1965
DB 361 TTGGCCAGGCTGTGCTGCAATCTCTGACCGCAAGTATCACCGGCTCTCGCTCCCAAA 302
QY 1966 GTGCTGGGATTAAGAGCGTAGAGCCAGCGGCGCCGCAAGGGGAACTCTTGAGGAG 2025
DB 301 GTGCTGGGATTAAGAGCGTAGAGCCAGCGGCGCCGCAAGGGGAACTCTTGAGGAG 242
QY 2026 GAGCAGAGGGGCTCACATCTCCCTCTGATTTCCCAATGACATTCCTTATCTCTCCC 2085
DB 241 GAGCAGAGGGGCTCACATCTCCCTCTGATTTCCCAATGACATTCCTTATCTCTCCC 182
QY 2086 ATCTACCCAGGATCTATTTGTCTTTCTTCTGCAATTTTACATATGATGTGATGTC 2145
DB 181 ATCTACCCAGGATCTATTTGTCTTTCTTCTGCAATTTTACATATGATGTGATGTC 122
QY 2146 GCTACACACACCCCGCCATGSGGGGATGAGAGGGGTGCAAGGCGCTGCTCTCCACT 2205
DB 121 GCTACACACACCCCGCCATGSGGGGATGAGAGGGGTGCAAGGCGCTGCTCTCCACT 62
QY 2206 TTTTCTACCTTGGAGCTGTATTAATTAATCACTTCTGTTTGTCAATTTTCA 2260
DB 61 TTTTCTACCTTGGAGCTGTATTAATTAATCACTTCTGTTTGTCAATTTTCA 7

RESULT 8
US-09-908-711-158/c
; Sequence 158, Application US/09908711
; Patent No. US20020045230A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA128
; CURRENT APPLICATION NUMBER: US/09/908,711
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US01/01360
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,867
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01344
; PRIOR FILING DATE: 2001-01-17
```

```
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01345
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,888
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01329
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,905
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01354
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,891
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01339
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01340
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,874
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01334
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,898
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01320
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,853
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01349
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,902
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01239
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,870
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01348
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,882
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01347
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,896
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01307
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,864
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01341
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,856
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01336
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,868
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/209,467
; PRIOR FILING DATE: 2000-06-07
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 158
; LENGTH: 32174
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

NAME/KEY: SITE	LOCATION: (29356)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29357)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29358)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29359)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29360)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29361)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29362)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29363)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29364)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29365)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29366)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29367)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29368)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29369)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29370)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29371)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29372)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29373)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29374)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29375)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29376)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29377)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29378)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE	LOCATION: (29379)	n equals a,t,g, or c
OTHER INFORMATION:		
NAME/KEY: SITE		

```

1 LOCATION: (29380)
2 OTHER INFORMATION: n equals a,t,g, or c
3 NAME/KEY: SITE
4 LOCATION: (29381)
5 OTHER INFORMATION: n equals a,t,g, or c
6 NAME/KEY: SITE
7 LOCATION: (29382)
8 OTHER INFORMATION: n equals a,t,g, or c
9 NAME/KEY: SITE
10 LOCATION: (29383)
11 OTHER INFORMATION: n equals a,t,g, or c
12 NAME/KEY: SITE
13 LOCATION: (29384)
14 OTHER INFORMATION: n equals a,t,g, or c
15 NAME/KEY: SITE
16 LOCATION: (29385)
17 OTHER INFORMATION: n equals a,t,g, or c
18 NAME/KEY: SITE
19 LOCATION: (29386)
20 OTHER INFORMATION: n equals a,t,g, or c
21 NAME/KEY: SITE
22 LOCATION: (29387)
23 OTHER INFORMATION: n equals a,t,g, or c
24 NAME/KEY: SITE
25 LOCATION: (29388)
26 OTHER INFORMATION: n equals a,t,g, or c
27 NAME/KEY: SITE
28 LOCATION: (29389)
29 OTHER INFORMATION: n equals a,t,g, or c
30 NAME/KEY: SITE
31 LOCATION: (29390)
32 OTHER INFORMATION: n equals a,t,g, or c
33 NAME/KEY: SITE
34 LOCATION: (29391)
35 OTHER INFORMATION: n equals a,t,g, or c
36 NAME/KEY: SITE
37 LOCATION: (29392)
38 OTHER INFORMATION: n equals a,t,g, or c
39 NAME/KEY: SITE
40 LOCATION: (29393)

```

Query Match	31.2%	Score 708.6;	DB 9;	Length 32174;
Best Local Similarity	99.4%;	Pred. No. 7.2e-209;		
Matches 711; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

Qy	1546	TTTCCCTGAGTGGCTGTGGCGGAACTGGAGCAGAAACTGTCGTCCCTTTTACAGAC	1605
Db	6050	TCTTGCAGAGTgacctgtgtgcggaaCTGTGTCAAAACTACTGTCTCCCTTTTACAGAC	5993
Qy	1606	TCCTTTGCCCCAGAGCAGAGATGAAAAAGCCAGAGGAGTGGAAAGATGCATGCTTCAGC	1665
Db	5990	TCCTTTGCCCCAGAGCAGAGAAATGAAAAAGCCAGAGGAGTGGAAAGATGCATGCTTCAGC	5933
Qy	1666	TGTGCTCTGTGCTGCTGCAGCCAAAGTCTTCAATTTGGGGCCAAAGGGGAAACTTTTTTGGAG	1725
Db	5930	TGTGCTCTGTGCTGCAGCCAAAGTCTTCAATTTGGGGCCAAAGGGGAAACTTTTTTGGAG	5873
Qy	1726	AAGCGCTCTTGCTTTGTACACCCAGCGTGGAAATGCAAGTGGCGGAACTCAGCTCAACGGAA	1789
Db	5870	AAGCGCTCTTGCTTTGTACACCCAGCGTGGAAATGCAAGTGGCGGAACTCAGCTCAACGGAA	5811
Qy	1786	CCTCCACCTCCTCGGGTTCAAGTGAATTTTCTGCTCAGCTCCGCCAAGTAGCTGGGAATAC	1845
Db	5810	CCTCCACCTCCTCGGGTTCAAGTGAATTTTCTGCTCAGCTCCGCCAAGTAGCTGGGAATAC	5755
Qy	1846	AGGCAACGCAACCATGCCCCAGCTAAATTTTGTATTTTCAATGAAACGGGAATTTCAACACG	1905
Db	5750	AGGCAACGCAACCATGCCCCAGCTAAATTTTGTATTTTCAATGAAACGGGAATTTCAACACG	5693
Qy	1906	TTGGCCAGAGCGTGTCTGGAATCCTGACCGCAAGTGAATCCACCGGCTCCGGCTCCCAAA	1965
Db	5690	TTGGCCAGAGCGTGTCTGGAATCCTGACCGCAAGTGAATCCACCGGCTCCGGCTCCCAAA	5633

1	OTHER INFORMATION:	n	equals	a,t,g,	or c
2	NAME/KEY: SITE				
3	LOCATION: (29392)				
4	OTHER INFORMATION:	n	equals	a,t,g,	or c
5	NAME/KEY: SITE				
6	LOCATION: (29393)				
7	OTHER INFORMATION:	n	equals	a,t,g,	or c
8	NAME/KEY: SITE				
9	LOCATION: (29394)				
10	OTHER INFORMATION:	n	equals	a,t,g,	or c
11	NAME/KEY: SITE				
12	LOCATION: (29395)				
13	OTHER INFORMATION:	n	equals	a,t,g,	or c
14	NAME/KEY: SITE				
15	LOCATION: (29396)				
16	OTHER INFORMATION:	n	equals	a,t,g,	or c
17	NAME/KEY: SITE				
18	LOCATION: (29397)				
19	OTHER INFORMATION:	n	equals	a,t,g,	or c
20	NAME/KEY: SITE				
21	LOCATION: (29398)				
22	OTHER INFORMATION:	n	equals	a,t,g,	or c
23	NAME/KEY: SITE				
24	LOCATION: (29399)				
25	OTHER INFORMATION:	n	equals	a,t,g,	or c
26	NAME/KEY: SITE				
27	LOCATION: (29400)				
28	OTHER INFORMATION:	n	equals	a,t,g,	or c
29	NAME/KEY: SITE				
30	LOCATION: (29401)				
31	OTHER INFORMATION:	n	equals	a,t,g,	or c
32	NAME/KEY: SITE				
33	LOCATION: (29402)				
34	OTHER INFORMATION:	n	equals	a,t,g,	or c
35	NAME/KEY: SITE				
36	LOCATION: (29403)				
37	OTHER INFORMATION:	n	equals	a,t,g,	or c
38	NAME/KEY: SITE				
39	LOCATION: (29404)				
40	OTHER INFORMATION:	n	equals	a,t,g,	or c
41	NAME/KEY: SITE				
42	LOCATION: (29405)				
43	OTHER INFORMATION:	n	equals	a,t,g,	or c
44	NAME/KEY: SITE				
45	LOCATION: (29406)				
46	OTHER INFORMATION:	n	equals	a,t,g,	or c
47	NAME/KEY: SITE				
48	LOCATION: (29407)				
49	OTHER INFORMATION:	n	equals	a,t,g,	or c
50	NAME/KEY: SITE				
51	LOCATION: (29408)				
52	OTHER INFORMATION:	n	equals	a,t,g,	or c
53	NAME/KEY: SITE				
54	LOCATION: (29409)				
55	OTHER INFORMATION:	n	equals	a,t,g,	or c
56	NAME/KEY: SITE				
57	LOCATION: (29410)				
58	OTHER INFORMATION:	n	equals	a,t,g,	or c
59	NAME/KEY: SITE				
60	LOCATION: (29411)				
61	OTHER INFORMATION:	n	equals	a,t,g,	or c
62	NAME/KEY: SITE				
63	LOCATION: (29412)				
64	OTHER INFORMATION:	n	equals	a,t,g,	or c
65	NAME/KEY: SITE				
66	LOCATION: (29413)				
67	OTHER INFORMATION:	n	equals	a,t,g,	or c
68	NAME/KEY: SITE				
69	LOCATION: (29414)				
70	OTHER INFORMATION:	n	equals	a,t,g,	or c
71	NAME/KEY: SITE				
72	LOCATION: (29415)				
73	OTHER INFORMATION:	n	equals	a,t,g,	or c

```

; NAME/KEY: SITE
; LOCATION: (29416)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE

```

Query Match	31.2%;	Score 708.6;	DB 9;	Length 32174;
Best Local Similarity	99.4%;	Pred. No. 7.2e-209;		
Matches 711;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;

Qy	1546	TTTTCCCTGGATGCGCGGGATCTGGATGCGAATACTAGTCCTCTTTCACAGAC	1605
Dp	6050	TCTTGAGAGTGGCTGTGCGGGATGGGTGCGAATACTAGTCCTCTTTCACAGAC	5991
Qy	1606	TTCTTTTCCCCAGAGCAGAGAAATGGAAAAGCCAGGAGGTGGAAATGAGATGCTTCCAGC	1666
Dp	5990	TTCTTTTCCCCAGAGCAGAGAAATGGAAAAGCCAGGAGGTGGAAATGAGATGAGTCTTCCAGC	5933
Qy	1666	TGTGCTCTGTGCTGCACACCAGCTTCTCATTTTGGGGCCAAAGGGGAAACTTTTTTTGGAG	1726
Dp	5930	TGTGCTCTGTGCTGCACACCAGCTTCTCATTTTGGGGCCAAAGGGGAAACTTTTTTTGGAG	5877
Qy	1726	AAGGCGTCTGTGCTTTGTTCACCCACGCTGGAAATGCAATGTCGGGATCTAGCTCACCGCA	1788
Dp	5870	AAGGCGTCTGTGCTTTGTTCACCCACGCTGGAAATGCAATGTCGGGATCTAGCTCACCGCA	5811
Qy	1786	CCTCACCCTCTGGGGTTCAAGTGAATTTTCCGCGCAGCCCTCCCAAGTAGCGGGATATC	1845
Dp	5810	CCTCACCCTCTGGGGTTCAAGTGAATTTTCCGCGCAGCCCTCCCAAGTAGCTGGGATATC	5755
Qy	1846	AGGCGACGCCAATGCGCCAGCTAAATTTTGTATTTTCAGTAGAAACGGGATTTTCCACAG	1905
Dp	5750	AGGCGACGCCAATGCGCCAGCTAAATTTTGTATTTTCAGTAGAAACGGGATTTTCCACAG	5691
Qy	1906	TTGGCCAGGCGTGTGCTGGAATCCTGTGACCGCAATGATCCACCGCGCTCCGCTCCCAA	1966
Dp	5690	TTGGCCAGGCGTGTGCTGGAATCCTGTGACCGCAATGATCCACCGCGCTCCGCTCCCAA	5631
Qy	1966	GTCGCTGGGATTTACAGCGGTGAGCACCAGCTGCGCGGCCCAAAAGGGGAAACTTTGTGGAG	2026
Dp	5630	GTCGCTGGGATTTACAGCGGTGAGCACCAGCTGCGCGGCCCAAAAGGGGAAACTTTGTGGAG	5577
Qy	2026	GAGCAGAGGGGCTCAATCTTCCCTCTGATTTCCCCATGCATTTGCCCTTATCTTCCCC	2086
Dp	5570	GAGCAGAGGGGCTCAATCTTCCCTCTGATTTCCCCATGCATTTGCCCTTATCTTCCCC	5511
Qy	2086	ATCTAGCCAGGAATCTATTTGTTTCTTCTGCAATTTCATGATTTGATGTGTC	2145
Dp	5510	ATCTAGCCAGGAATCTATTTGTTTCTTCTGCAATTTCATGATTTGATGTGTC	5451
Qy	2146	GCTACCAACACCCCCCATGGGGGGGTGAGAGGGGTGCAAGGCCCTGCTGCACCT	2205
Dp	5450	GCTACCAACACCCCCCATGGGGGGGTGAGAGGGGTGCAAGGCCCTGCTGCACCT	5391
Qy	2206	TTTTCTACCTTTGGAACGTATTAATAAATCACTTCTGTTTGTTCAGTTTTTCA	2260
Dp	5390	TTTTCTACCTTTGGAACGTATTAATAAATCACTTCTGTTTGTTCAGTTTTTCA	5336

```

RESULT 10
US-09-764-877-2645/c
; Sequence 2645, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2645
; LENGTH: 32174

```


[illegible]

Qy	1966	GTGCTGGGATTTACAGGGGGTGAGACCGGTGCGGGGCCCAAGGGGAAACCTTTGTGGAG	2025
Db	5630	GTGCTGGGATTTACAGGGGGTGAGACCGGTGCGGGGCCCAAGGGGAAACCTTTGTGGAG	5571
Qy	2026	GAGCAGAGGGGGCTCACATCTTCCCTCTGATTTCCCAATGACATTTGCTTCTCTCCC	2085
Db	5570	GAGCAGAGGGGGCTCACATCTTCCCTCTGATTTCCCAATGACATTTGCTTCTCTCCC	5511
Qy	2086	ATCTAGCCAGGAATCTATTTGTTTCTTTCTTCTGCAATTACTATGATTTGTATGTGCC	2145
Db	5510	ATCTAGCCAGGAATCTATTTGTTTCTTTCTTCTGCAATTACTATGATTTGTATGTGCC	5451
Qy	2146	GCTACCAACACACCCCCCATGGGGGGGGGAGAGAGGGGGTCAAGAGCCCTGCTCCACT	2205
Db	5450	GCTACCAACACACCCCCCATGGGGGGGGGAGAGAGGGGGTCAAGAGCCCTGCTCCACT	5391
Qy	2206	TTTTCTACTCTTGAACCTGTATTAGATAAATCACTTCTGTTTGTTCAGTTTTTCA	2260
Db	5390	TTTTCTACTCTTGAACCTGTATTAGATAAATCACTTCTGTTTGTTCAGTTTTTCA	5336
RESULT 11			
US-09-860-670-232/c			
Sequence 232, Application us/09860670			
Patent No. US20020165137A1			
GENERAL INFORMATION:			
APPLICANT: Ruben et al.			
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies			
FILE REFERENCE: PA127P1			
CURRENT APPLICATION NUMBER: US/09/860,670			
CURRENT FILING DATE: 2001-05-21			
Prior application data removed - consult PALM or file wrapper			
NUMBER OF SEQ ID NOS: 289			
SOFTWARE: PatentIn Ver. 2.0			
SEQ ID NO 232			
LENGTH: 32174			
TYPE: DNA			
ORGANISM: Homo sapiens			
FEATURE:			
NAME/KEY: SITE			
LOCATION: (29356)			
OTHER INFORMATION: n equals a,t,g, or c			
NAME/KEY: SITE			
LOCATION: (29357)			
OTHER INFORMATION: n equals a,t,g, or c			
NAME/KEY: SITE			
LOCATION: (29358)			
OTHER INFORMATION: n equals a,t,g, or c			
NAME/KEY: SITE			
LOCATION: (29359)			
OTHER INFORMATION: n equals a,t,g, or c			
NAME/KEY: SITE			
LOCATION: (29360)			
OTHER INFORMATION: n equals a,t,g, or c			
NAME/KEY: SITE			
LOCATION: (29361)			
OTHER INFORMATION: n equals a,t,g, or c			
NAME/KEY: SITE			
LOCATION: (29362)			
OTHER INFORMATION: n equals a,t,g, or c			
NAME/KEY: SITE			
LOCATION: (29363)			
OTHER INFORMATION: n equals a,t,g, or c			
NAME/KEY: SITE			
LOCATION: (29364)			
OTHER INFORMATION: n equals a,t,g, or c			
NAME/KEY: SITE			
LOCATION: (29365)			
OTHER INFORMATION: n equals a,t,g, or c			
NAME/KEY: SITE			
LOCATION: (29366)			
OTHER INFORMATION: n equals a,t,g, or c			

1	LOCATION: (29391)	n	equals	a,t,g,	or c
2	OTHER INFORMATION:	n	equals	a,t,g,	or c
3	NAME/KEY: SITE	n	equals	a,t,g,	or c
4	LOCATION: (29392)	n	equals	a,t,g,	or c
5	OTHER INFORMATION:	n	equals	a,t,g,	or c
6	NAME/KEY: SITE	n	equals	a,t,g,	or c
7	LOCATION: (29393)	n	equals	a,t,g,	or c
8	OTHER INFORMATION:	n	equals	a,t,g,	or c
9	NAME/KEY: SITE	n	equals	a,t,g,	or c
10	LOCATION: (29394)	n	equals	a,t,g,	or c
11	OTHER INFORMATION:	n	equals	a,t,g,	or c
12	NAME/KEY: SITE	n	equals	a,t,g,	or c
13	LOCATION: (29395)	n	equals	a,t,g,	or c
14	OTHER INFORMATION:	n	equals	a,t,g,	or c
15	NAME/KEY: SITE	n	equals	a,t,g,	or c
16	LOCATION: (29396)	n	equals	a,t,g,	or c
17	OTHER INFORMATION:	n	equals	a,t,g,	or c
18	NAME/KEY: SITE	n	equals	a,t,g,	or c
19	LOCATION: (29397)	n	equals	a,t,g,	or c
20	OTHER INFORMATION:	n	equals	a,t,g,	or c
21	NAME/KEY: SITE	n	equals	a,t,g,	or c
22	LOCATION: (29398)	n	equals	a,t,g,	or c
23	OTHER INFORMATION:	n	equals	a,t,g,	or c
24	NAME/KEY: SITE	n	equals	a,t,g,	or c
25	LOCATION: (29399)	n	equals	a,t,g,	or c
26	OTHER INFORMATION:	n	equals	a,t,g,	or c
27	NAME/KEY: SITE	n	equals	a,t,g,	or c
28	LOCATION: (29400)	n	equals	a,t,g,	or c
29	OTHER INFORMATION:	n	equals	a,t,g,	or c
30	NAME/KEY: SITE	n	equals	a,t,g,	or c
31	LOCATION: (29401)	n	equals	a,t,g,	or c
32	OTHER INFORMATION:	n	equals	a,t,g,	or c
33	NAME/KEY: SITE	n	equals	a,t,g,	or c
34	LOCATION: (29402)	n	equals	a,t,g,	or c
35	OTHER INFORMATION:	n	equals	a,t,g,	or c
36	NAME/KEY: SITE	n	equals	a,t,g,	or c
37	LOCATION: (29403)	n	equals	a,t,g,	or c
38	OTHER INFORMATION:	n	equals	a,t,g,	or c
39	NAME/KEY: SITE	n	equals	a,t,g,	or c
40	LOCATION: (29404)	n	equals	a,t,g,	or c
41	OTHER INFORMATION:	n	equals	a,t,g,	or c
42	NAME/KEY: SITE	n	equals	a,t,g,	or c
43	LOCATION: (29405)	n	equals	a,t,g,	or c
44	OTHER INFORMATION:	n	equals	a,t,g,	or c
45	NAME/KEY: SITE	n	equals	a,t,g,	or c
46	LOCATION: (29406)	n	equals	a,t,g,	or c
47	OTHER INFORMATION:	n	equals	a,t,g,	or c
48	NAME/KEY: SITE	n	equals	a,t,g,	or c
49	LOCATION: (29407)	n	equals	a,t,g,	or c
50	OTHER INFORMATION:	n	equals	a,t,g,	or c
51	NAME/KEY: SITE	n	equals	a,t,g,	or c
52	LOCATION: (29408)	n	equals	a,t,g,	or c
53	OTHER INFORMATION:	n	equals	a,t,g,	or c
54	NAME/KEY: SITE	n	equals	a,t,g,	or c
55	LOCATION: (29409)	n	equals	a,t,g,	or c
56	OTHER INFORMATION:	n	equals	a,t,g,	or c
57	NAME/KEY: SITE	n	equals	a,t,g,	or c
58	LOCATION: (29410)	n	equals	a,t,g,	or c
59	OTHER INFORMATION:	n	equals	a,t,g,	or c
60	NAME/KEY: SITE	n	equals	a,t,g,	or c
61	LOCATION: (29411)	n	equals	a,t,g,	or c
62	OTHER INFORMATION:	n	equals	a,t,g,	or c
63	NAME/KEY: SITE	n	equals	a,t,g,	or c
64	LOCATION: (29412)	n	equals	a,t,g,	or c
65	OTHER INFORMATION:	n	equals	a,t,g,	or c
66	NAME/KEY: SITE	n	equals	a,t,g,	or c
67	LOCATION: (29413)	n	equals	a,t,g,	or c
68	OTHER INFORMATION:	n	equals	a,t,g,	or c
69	NAME/KEY: SITE	n	equals	a,t,g,	or c
70	LOCATION: (29414)	n	equals	a,t,g,	or c
71	OTHER INFORMATION:	n	equals	a,t,g,	or c
72	NAME/KEY: SITE	n	equals	a,t,g,	or c
73	LOCATION: (29415)	n	equals	a,t,g,	or c


```
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: SITE
/ LOCATION: (29416)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: SITE

Query Match      31.2%; Score 708.6; DB 10; Length 32174;
Best Local Similarity 99.4%; Pred. No. 7.2e-209;
Matches 711; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1546 TTTCCCTGGTGGCTGTGGGGGAGCTGTGCAGAACTACTGCTCTCTTTTTCACAGCAC 1605
DB 6050 TCTTGAGGTGGCTGTGGGGGAGCTGTGCAGAACTACTGCTCTCTTTTTCACAGCAC 5991
QY 1606 TCTTTGGCCCCAGAGAGAAATGAGAAAGCCAGGAGGAGAGTGCATGCTTCACAC 1665
DB 5990 TCTTTGGCCCCAGAGAGAAATGAGAAAGCCAGGAGGAGAGTGCATGCTTCACAC 5931
QY 1666 TGTGCTCTGTGTCAGCCAGCAAGTCTTCATTTGGGGCCAAAGGGGAACTTTTGTGGAG 1725
DB 5930 TGTGCTCTGTGTCAGCCAGCAAGTCTTCATTTGGGGCCAAAGGGGAACTTTTGTGGAG 5871
QY 1726 AAGGCTCTTGTCTTTGTACCCAGCTGGAATGAGTGGGGGATCTCAGCTCACCGCAA 1785
DB 5870 AAGGCTCTTGTCTTTGTACCCAGCTGGAATGAGTGGGGGATCTCAGCTCACCGCAA 5811
QY 1786 CCTCCACCTCTGGGGTTCAAGATTTTCTGCTCAGCTCCCACTAGCTGGGAATAC 1845
DB 5810 CTTCCACCTCTGGGGTTCAAGATTTTCTGCTCAGCTCCCACTAGCTGGGAATAC 5751
QY 1846 AAGCAAGCCAGCAATGCCAGCAATTTTGTATTTTTCAGTAAAGGAGATTTTCACACAG 1905
DB 5750 AAGCAAGCCAGCAATGCCAGCAATTTTGTATTTTTCAGTAAAGGAGATTTTCACACAG 5691
QY 1906 TTGGCCAGGCTGTCTTGAATCTCTTACCGCAAGTATTCACCCGCTCTGCCCAA 1965
DB 5690 TTGGCCAGGCTGTCTTGAATCTCTTACCGCAAGTATTCACCCGCTCTGCCCAA 5631
QY 1966 GTGCTGGGATTACAGGCGGAGAGCCAGCCCGGCCCAAGGGGAACTTTGTGGGAG 2025
DB 5630 GTGCTGGGATTACAGGCGGAGAGCCAGCCCGGCCCAAGGGGAACTTTGTGGGAG 5571
QY 2026 GAGCAGAGGGGCTACATCTCCCTCTGATTTCCCATGCACTTCTCTCTCC 2085
DB 5570 GAGCAGAGGGGCTACATCTCCCTCTGATTTCCCATGCACTTCTCTCTCC 5511
QY 2086 ATCTAGCCAGGAATCTATTGTGTTTCTTCTGCAATTTACTATGATGTGATGTC 2145
DB 5510 ATCTAGCCAGGAATCTATTGTGTTTCTTCTGCAATTTACTATGATGTGATGTC 5451
QY 2146 GGTACCAACACCCCCCATTTGGGGGGTGGAGAGGGGTGCAAGCCCTGCTCTCACT 2205
DB 5450 GGTACCAACACCCCCCATTTGGGGGGTGGAGAGGGGTGCAAGCCCTGCTCTCACT 5391
QY 2206 TTTTCTACCTTGAATCTATTAGATAAATCACTCTGTTGTTGTCAGTTTTTCA 2260
DB 5390 TTTTCTACCTTGAATCTATTAGATAAATCACTCTGTTGTTGTCAGTTTTTCA 5336

RESULT 12
US-09-764-904-90/c.
; Sequence 90, Application US/09764904
; Patent No. US20020173454A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA122
; CURRENT APPLICATION NUMBER: US/09/764,904
; PRIORITY FILING DATE: 2001-01-17
; PRIOR APPLICATION data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 90
```

```
LENGTH: 32174
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (29356)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29358)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29357)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29359)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29360)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29361)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29362)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29363)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29364)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29365)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29366)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29367)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29368)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29369)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29370)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29371)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29372)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29373)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29374)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29375)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29376)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29377)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29378)
OTHER INFORMATION: n equals a,t,g, or c
```

```
NAME/KEY: SITE
LOCATION: (29379)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29380)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29381)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29382)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29383)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29384)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29385)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29386)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29387)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29388)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29389)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29390)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29391)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29392)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29393)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29394)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29395)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29396)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29397)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29398)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29399)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29400)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29401)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29402)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
```

```
LOCATION: (29403)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29404)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29405)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29406)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29407)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29408)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29409)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29410)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29411)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29412)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29413)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29414)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29415)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29416)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
```

Query Match 31.2%; Score 708.6; DB 10; Length 32174;
Best Local Similarity 99.4%; Pred. No. 7.2e-209;
Matches 711; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 1546 TTTCCTGTGGCTGTGGGAGCTGTGCAGAACTACTGCTTTCACAGCAC 1605
DB 6050 TCTTGAGGTGGCTGTGGGAGCTGTGCAGAACTACTGCTTTCACAGCAC 5991
QY 1606 TCCTTTGCCCGAGAGAGATGAAAAGCCAGGAGGTGAAAGTCATGCTTCACG 1665
DB 5990 TCTTTGCCCGAGAGAGATGAAAAGCCAGGAGGTGAAAGTCATGCTTCACG 5931
QY 1666 TGTGCTGTGGCTGTGGGAGCTGTGCAGAACTACTGCTTTCACAGCAC 1725
DB 5930 TGTGCTGTGGCTGTGGGAGCTGTGCAGAACTACTGCTTTCACAGCAC 5871
QY 1726 AAGGCGTGTGGCTTTCACAGCTGTGCAGAACTACTGCTTTCACAGCAC 1785
DB 5870 AAGGCGTGTGGCTTTCACAGCTGTGCAGAACTACTGCTTTCACAGCAC 5811
QY 1786 CTTGCACCTCTGTGGTTCAAGTATTTCTGCTTCAGCTCCCAAGTATGCGAAATAC 1845
DB 5810 CTTGCACCTCTGTGGTTCAAGTATTTCTGCTTCAGCTCCCAAGTATGCGAAATAC 5751
QY 1846 AAGCAGCCACCATGCTGCTCAATTTTGTATTTTCAATAGAAAGGATTTACACAG 1905
DB 5750 AAGCAGCCACCATGCTGCTCAATTTTGTATTTTCAATAGAAAGGATTTACACAG 5691
QY 1906 TTGGCAGGCTGTGCTGCAACTCTGACCGCAAGTATCCACCGCTCGCTCCCAAA 1965
```

Db 5690 TTGGCCAGGCTGCTCGAATCTCGAATCCGCAAGTATCCACCCGCTCCGCTCCCAAA 5631
Qy 1966 GTGCTGGATTACAGGCGTGAAGCCACCGTGGCCGCGCCAAAGGGGAACTCTTGAGGAG 2025
Db 5630 GTGCTGGATTACAGGCGTGAAGCCACCGTGGCCGCGCCAAAGGGGAACTCTTGAGGAG 5571
Qy 2026 GAGCAGAGGGGCTCACATCTCCCTCTGATTCCCGCATGTCATTCCTTATCTCTCCC 2085
Db 5570 GAGCAGAGGGGCTCACATCTCCCTCTGATTCCCGCATGTCATTCCTTATCTCTCCC 5511
Qy 2086 ATCTAGCCAGATCTATTGTTGTTTCTTCTGCAATTTACTATATGTTGATGTGCC 2145
Db 5510 ATCTAGCCAGATCTATTGTTGTTTCTTCTGCAATTTACTATATGTTGATGTGCC 5451
Qy 2146 GCTACCAACCCCGCCCATGGGGGGGTGAGAGGGGGTCAAGGGCCCTGCTGCTCACT 2205
Db 5450 GCTACCAACCCCGCCCATGGGGGGGTGAGAGGGGGTCAAGGGCCCTGCTGCTCACT 5391
Qy 2206 TTTTCTACCTTGGAACTGTATTAGATMAATCATTCTGTTGTCAGTTTTTCA 2260
Db 5390 TTTTCTACCTTGGAACTGTATTAGATMAATCATTCTGTTGTCAGTTTTTCA 5336

RESULT 13

US-09-764-891-6480/c
Sequence 6480, Application US/09764891

Publication No. US20030077808A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC006

CURRENT APPLICATION NUMBER: US/09/764,891

CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 10231

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 6480

LENGTH: 32174

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (29356)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29357)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29358)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29359)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29360)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29361)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29362)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29363)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29364)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29365)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29366)

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29367)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29368)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29369)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29370)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29371)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29372)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29373)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29374)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29375)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29376)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29377)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29378)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29379)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29380)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29381)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29382)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29383)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29384)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29385)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29386)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29387)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29388)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29389)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29390)
OTHER INFORMATION: n equals a,t,g, or c

```
; NAME/KEY: SITE
; LOCATION: (29391)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29392)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29393)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29394)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29395)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29396)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29397)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29398)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29399)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29400)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29401)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29402)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29403)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29404)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29405)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29406)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29407)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29408)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29409)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29410)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29411)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29412)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29413)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29414)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
```

```
; LOCATION: (29415)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29416)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
```

Query Match 31.2%; Score 708.6; DB 11; Length 32174;
Best Local Similarity 99.4%; Pred. No. 7.2e-209;
Matches 711; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 1546 TTTCCCTGTGGCTGTGGGAGGAGTGCAGAAACTACTCTCTCCCTTTTACAGCAGC 1605
DB 6050 TCTTGAGGTGGCTGTGGGAGGAGTGCAGAAACTACTCTCTCTTTTACAGCAGC 5991
QY 1606 TCTTTTGGCCAGAGAGAGAAATGAAAAGCCAGGAGGTGAGATGATGATCTTCAGC 1665
DB 5990 TCTTTTGGCCAGAGAGAGAAATGAAAAGCCAGGAGGTGAGATGATGATCTTCAGC 5931
QY 1666 TGTGCTCTGTGCTGCGCAGCAAGTCTTCTATTTTGGGGCCAAAGGGAACTTTTGTGAG 1725
DB 5930 TGTGCTCTGTGCTGCGCAGCAAGTCTTCTATTTTGGGGCCAAAGGGAACTTTTGTGAG 5871
QY 1726 AAGGCGCTTGTGCTTGTGCAACCGCTGAGATGAGTGGCGGATCTCAGTCCAGCAG 1785
DB 5870 AAGGCGCTTGTGCTTGTGCAACCGCTGAGATGAGTGGCGGATCTCAGTCCAGCAG 5811
QY 1786 CCTCCACTCTGTGGGTTCAAGTATTTTCTGCTTCAAGCTTCCCAAGTATGTTGGAGATAC 1845
DB 5810 CCTCCACTCTGTGGGTTCAAGTATTTTCTGCTTCAAGCTTCCCAAGTATGTTGGAGATAC 5751
QY 1846 AAGCAGCCAGCATGCGCCAGCTAATTTTGTATTTTCAATAGTAAACGGATTTCCAGCAG 1905
DB 5750 AAGCAGCCAGCATGCGCCAGCTAATTTTGTATTTTCAATAGTAAACGGATTTCCAGCAG 5691
QY 1906 TTGGCAGGCTGATCTGCAACTCTGACCGCAAGTATCACCCTCCGCTCCCAAA 1965
DB 5690 TTGGCAGGCTGATCTGCAACTCTGACCGCAAGTATCACCCTCCGCTCCCAAA 5631
QY 1966 GTGCTGGATTAACAGGCTGAGCCACCGTCCCGGCCAAAGGGGAACTTTTGTGGAG 2025
DB 5630 GTGCTGGATTAACAGGCTGAGCCACCGTCCCGGCCAAAGGGGAACTTTTGTGGAG 5571
QY 2026 GAGCAGAGGGCTCATATCTCCCTCTGATCTCCCGCAGCATGCTTATCTCTCC 2085
DB 5570 GAGCAGAGGGCTCATATCTCCCTCTGATCTCCCGCAGCATGCTTATCTCTCC 5511
QY 2086 ATCTAGCCAGATCTATTTGTTTTCTTCTGCAATTTACTATGATGTATGTGCC 2145
DB 5510 ATCTAGCCAGATCTATTTGTTTTCTTCTGCAATTTACTATGATGTATGTGCC 5451
QY 2146 GCTACCAACACCCCCCATGAGGGGGTGAAGAGGGGTGCAAGCCCTGCTCCACT 2205
DB 5450 GCTACCAACACCCCCCATGAGGGGGTGAAGAGGGGTGCAAGCCCTGCTCCACT 5391
QY 2206 TTTTCTACCTGGAATCTGATTAATTAATCACTCTGTTGTCAGTTTCA 2260
DB 5390 TTTTCTACCTGGAATCTGATTAATTAATCACTCTGTTGTCAGTTTCA 5336
```

RESULT 14
US-09-764-891-10135
; Sequence 10135, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; PRIOR APPLICATION DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0

1	SEO ID NO 10135
2	LENGTH: 32174
3	TYPE: DNA
4	ORGANISM: Homo sapiens
5	FEATURE:
6	NAME/KEY: SITE
7	LOCATION: (1295)
8	OTHER INFORMATION: n equals a,t,g, or c
9	NAME/KEY: SITE
10	LOCATION: (1296)
11	OTHER INFORMATION: n equals a,t,g, or c
12	NAME/KEY: SITE
13	LOCATION: (1297)
14	OTHER INFORMATION: n equals a,t,g, or c
15	NAME/KEY: SITE
16	LOCATION: (1298)
17	OTHER INFORMATION: n equals a,t,g, or c
18	NAME/KEY: SITE
19	LOCATION: (1289)
20	OTHER INFORMATION: n equals a,t,g, or c
21	NAME/KEY: SITE
22	LOCATION: (1300)
23	OTHER INFORMATION: n equals a,t,g, or c
24	NAME/KEY: SITE
25	LOCATION: (1301)
26	OTHER INFORMATION: n equals a,t,g, or c
27	NAME/KEY: SITE
28	LOCATION: (1302)
29	OTHER INFORMATION: n equals a,t,g, or c
30	NAME/KEY: SITE
31	LOCATION: (1303)
32	OTHER INFORMATION: n equals a,t,g, or c
33	NAME/KEY: SITE
34	LOCATION: (1304)
35	OTHER INFORMATION: n equals a,t,g, or c
36	NAME/KEY: SITE
37	LOCATION: (1305)
38	OTHER INFORMATION: n equals a,t,g, or c
39	NAME/KEY: SITE
40	LOCATION: (1306)
41	OTHER INFORMATION: n equals a,t,g, or c
42	NAME/KEY: SITE
43	LOCATION: (1307)
44	OTHER INFORMATION: n equals a,t,g, or c
45	NAME/KEY: SITE
46	LOCATION: (1308)
47	OTHER INFORMATION: n equals a,t,g, or c
48	NAME/KEY: SITE
49	LOCATION: (1309)
50	OTHER INFORMATION: n equals a,t,g, or c
51	NAME/KEY: SITE
52	LOCATION: (1310)
53	OTHER INFORMATION: n equals a,t,g, or c
54	NAME/KEY: SITE
55	LOCATION: (1311)
56	OTHER INFORMATION: n equals a,t,g, or c
57	NAME/KEY: SITE
58	LOCATION: (1312)
59	OTHER INFORMATION: n equals a,t,g, or c
60	NAME/KEY: SITE
61	LOCATION: (1313)
62	OTHER INFORMATION: n equals a,t,g, or c
63	NAME/KEY: SITE
64	LOCATION: (1314)
65	OTHER INFORMATION: n equals a,t,g, or c
66	NAME/KEY: SITE
67	LOCATION: (1315)
68	OTHER INFORMATION: n equals a,t,g, or c
69	NAME/KEY: SITE
70	LOCATION: (1316)
71	OTHER INFORMATION: n equals a,t,g, or c
72	NAME/KEY: SITE
73	LOCATION: (1317)
74	OTHER INFORMATION: n equals a,t,g, or c
75	NAME/KEY: SITE
76	LOCATION: (1318)
77	OTHER INFORMATION: n equals a,t,g, or c
78	NAME/KEY: SITE
79	LOCATION: (1319)
80	OTHER INFORMATION: n equals a,t,g, or c
81	NAME/KEY: SITE
82	LOCATION: (1320)
83	OTHER INFORMATION: n equals a,t,g, or c
84	NAME/KEY: SITE
85	LOCATION: (1321)
86	OTHER INFORMATION: n equals a,t,g, or c
87	NAME/KEY: SITE
88	LOCATION: (1322)
89	OTHER INFORMATION: n equals a,t,g, or c
90	NAME/KEY: SITE
91	LOCATION: (1323)
92	OTHER INFORMATION: n equals a,t,g, or c
93	NAME/KEY: SITE
94	LOCATION: (1324)
95	OTHER INFORMATION: n equals a,t,g, or c
96	NAME/KEY: SITE
97	LOCATION: (1325)
98	OTHER INFORMATION: n equals a,t,g, or c
99	NAME/KEY: SITE
100	LOCATION: (1326)
101	OTHER INFORMATION: n equals a,t,g, or c
102	NAME/KEY: SITE
103	LOCATION: (1327)
104	OTHER INFORMATION: n equals a,t,g, or c
105	NAME/KEY: SITE
106	LOCATION: (1328)
107	OTHER INFORMATION: n equals a,t,g, or c
108	NAME/KEY: SITE
109	LOCATION: (1329)
110	OTHER INFORMATION: n equals a,t,g, or c
111	NAME/KEY: SITE
112	LOCATION: (1330)
113	OTHER INFORMATION: n equals a,t,g, or c
114	NAME/KEY: SITE
115	LOCATION: (1331)
116	OTHER INFORMATION: n equals a,t,g, or c
117	NAME/KEY: SITE
118	LOCATION: (1332)
119	OTHER INFORMATION: n equals a,t,g, or c
120	NAME/KEY: SITE
121	LOCATION: (1333)
122	OTHER INFORMATION: n equals a,t,g, or c
123	NAME/KEY: SITE
124	LOCATION: (1334)
125	OTHER INFORMATION: n equals a,t,g, or c
126	NAME/KEY: SITE
127	LOCATION: (1335)
128	OTHER INFORMATION: n equals a,t,g, or c
129	NAME/KEY: SITE
130	LOCATION: (1336)
131	OTHER INFORMATION: n equals a,t,g, or c
132	NAME/KEY: SITE
133	LOCATION: (1337)
134	OTHER INFORMATION: n equals a,t,g, or c
135	NAME/KEY: SITE
136	LOCATION: (1338)
137	OTHER INFORMATION: n equals a,t,g, or c
138	NAME/KEY: SITE
139	LOCATION: (1339)
140	OTHER INFORMATION: n equals a,t,g, or c
141	NAME/KEY: SITE
142	LOCATION: (1340)
143	OTHER INFORMATION: n equals a,t,g, or c
144	NAME/KEY: SITE
145	LOCATION: (1341)
146	OTHER INFORMATION: n equals a,t,g, or c

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: November 22, 2003, 02:28:33 ; Search time 135 seconds

(without alignments)
7431.584 Million cell updates/sec

Title: US-09-938-803-26

Perfect score: 2273
Sequence: 1 959999tgaagcgtacgtc.....ttttcaaaaaaaaaaaaaa 2273

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	223.4	9.8	1043	US-09-165-868-4	Sequence 4, Appli
C 2	223.4	9.8	1601	US-09-016-434-1218	Sequence 1218, Ap
C 3	221.8	9.8	90541	US-09-759-359A-3	Sequence 3, Appli
C 4	221.4	9.7	15977	US-09-608-285A-59	Sequence 59, Appli
C 5	218.8	9.6	14747	US-09-608-285A-42	Sequence 42, Appli
C 6	218.8	9.6	14747	US-09-557-800C-42	Sequence 42, Appli
C 7	218.6	9.6	62804	US-09-800-960-3	Sequence 3, Appli
C 8	218.4	9.6	283	US-08-579-445-26	Sequence 26, Appli
C 9	218	9.6	282	US-08-133-629-8	Sequence 8, Appli
C 10	217.8	9.6	14581	US-08-520-373D-4	Sequence 4, Appli
C 11	217.8	9.6	22481	US-08-367-841A-43	Sequence 43, Appli
C 12	217.8	9.6	22481	PCT-US95-07201-43	Sequence 43, Appli
C 13	217.8	9.6	22484	US-09-875-223-2	Sequence 2, Appli
C 14	215.6	9.5	21784	US-09-820-002-3	Sequence 3, Appli
C 15	215.6	9.5	42571	US-09-810-347-3	Sequence 3, Appli
C 16	215.4	9.5	174493	US-09-804-471A-3	Sequence 3, Appli
C 17	215	9.5	29332	US-09-016-434-1419	Sequence 1419, Ap
C 18	214.6	9.4	20303	US-08-370-975B-6	Sequence 6, Appli
C 19	214.6	9.4	26764	US-07-906-871-15	Sequence 15, Appli
C 20	214.4	9.4	17327	US-08-068-945A-1	Sequence 1, Appli
C 21	213.8	9.4	11531	US-08-442-806-1	Sequence 1, Appli
C 22	213.8	9.4	11531	US-09-355-295B-1	Sequence 3, Appli
C 23	213.8	9.4	11531	US-09-735-934A-3	Sequence 3, Appli
C 24	213.6	9.4	43950	US-10-060-332-3	Sequence 3, Appli
C 25	213.6	9.4	43950	US-09-801-861-3	Sequence 3, Appli
C 26	213	9.4	53332	US-08-781-891-79	Sequence 79, Appli
C 27	213	9.4	87350		

C 28	213	9.4	87350	US-09-618-166-79	Sequence 79, Appli
C 29	213	9.4	87543	US-09-791-211-3	Sequence 3, Appli
C 30	212.8	9.4	36741	US-09-301-665-3	Sequence 3, Appli
C 31	212.6	9.4	72604	US-09-268-992-7	Sequence 7, Appli
C 32	212.6	9.4	72604	US-09-657-474-7	Sequence 7, Appli
C 33	212.4	9.3	2758	US-09-620-312D-884	Sequence 884, App
C 34	212.2	9.3	55827	US-09-813-133A-3	Sequence 3, Appli
C 35	212.2	9.3	111282	US-09-754-250-3	Sequence 3, Appli
C 36	212	9.3	1988	US-08-257-963B-11	Sequence 11, Appli
C 37	212	9.3	1988	US-08-367-841A-11	Sequence 11, Appli
C 38	212	9.3	1988	PCT-US95-07201-11	Sequence 11, Appli
C 39	212	9.3	2000	US-09-705-267A-19	Sequence 19, Appli
C 40	212	9.3	5262	US-08-520-373D-5	Sequence 5, Appli
C 41	212	9.3	80246	US-09-078-294-4	Sequence 4, Appli
C 42	212	9.3	80595	US-09-078-294-4	Sequence 3, Appli
C 43	211.4	9.3	631	US-09-385-982-354	Sequence 354, App
C 44	211.4	9.3	66804	US-09-740-041-3	Sequence 3, Appli
C 45	211.2	9.3	50000	US-09-146-053-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-09-165-868-4/c
; Sequence 4, Application US/09165868
; Patent No. 6465176
; GENERAL INFORMATION:
; APPLICANT: Message Pharmaceuticals, Inc.
; TITLE OF INVENTION: METHOD FOR IDENTIFYING COMPOUNDS
; FILE REFERENCE: 50093/003W01
; CURRENT FILING DATE: 1999-10-02
; PRIOR APPLICATION NUMBER: US/09/165,868
; PRIOR FILING DATE: 1999-09-16
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1043
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-165-868-4

Query Match          9.8%; Score 223.4; DB 4; Length 1043;
Best Local Similarity 87.1%; Pred. No. 1.2e-51;
Matches 257; Conservative 0; Mismatches 36; Indels 2; Gaps 1;

QY 1714 TTTTGGGGAAGGCGCTTGTTCACCCGCGTGAATGCGGCGGATCTC 1773
    |||||
DB 906 TATTTTGGAGACGAGCTCTGTCTACCCGCGTGAATGCGGCGGATATTC 847

QY 1774 AGCTACCGCACTCTCACTCTCGGTTCAAGTATTTCTCGCTCAAGCT 1833
    |||||
DB 846 AGCTACGCACTCTCACTCTCGGTTCAAGTATTTCTCGCTCAAGCT 787

QY 1834 AGCTGGAAATACAG-GCAGCGCACATGCCAGTAAATTTTGTATTTCAAGTAAAC 1891
    |||||
DB 786 AGCTGGAAATACAGGCGCGCACATGCCAGTAAATTTTGTATTTTGTAGAGAC 727

QY 1892 GGGATTACACAGCTGGCGAGCGTCTGAACTCTGACCGGAATGCCACCGGC 1951
    |||||
DB 726 GGGATTACACAGCTGGCGAGCGTCTGAACTCTGACCTGAAGTATTCACCGGC 667

QY 1952 CTCGCGCTCCCAAGATGCTGGATTCAGAGGCGTGAACCGCGCCGCAAA 2006
    |||||
DB 666 CTCACCTCCCAAGATGCTGGATTCAGAGGCGTGAACCGCGCCGCGCTAGA 612

RESULT 2
US-09-016-434-1218/c
```

Sequence 1218, Application US/09016434
Patent No. 6500938
GENERAL INFORMATION:
APPLICANT: Janice Au-Young
APPLICANT: Jeffrey J. Seilhamer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
TITLE OF INVENTION: PATHWAY GENE EXPRESSION
NUMBER OF SEQUENCES: 1490
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,434
FILING DATE: HERewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0002 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1218:
SEQUENCE CHARACTERISTICS:
LENGTH: 1601 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENEBANK
CLONE: 9186270
US-09-016-434-1218

[illegible]

```

Sequence 3 Application US/09759359A
Patent No. 6492153
GENERAL INFORMATION:
APPLICANT: ABU-THREIDEH, Jane et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: C001043
CURRENT APPLICATION NUMBER: US/09/759,359A
CURRENT FILING DATE: 2001-01-16
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 90541
TYPE: DNA
ORGANISM: Human
US-09-759-359A-3

```

Query Match	Best Local Similarity	9.8%	Score 221.8	DB 4	Length 90541
Matches 247	Conservative	0	Mismatches 42	Indels	0
		0	Gaps		0
Qy	1214	TTTTTTTGGAAGAGGGCTTGTCTTTGTTCACCCACGCGTGAATGAGTGGCGGATCTC	1773		
Db	80683	TTTTTTTTAGATGAGAGTCTTCTCTGTTCGCCAGGCTGAGTGAAGTGGCGCAATCTT	80742		
Qy	1774	AGCTCACCGCAACTCCACTCTCTGGGTTCAGTATTTTCTCGCTCAGCTCCCAAGT	1833		
Db	80743	AGCTCACTGCACACCTGTGCTCCCAAGTTCACCAATTTCTCGCCGACGCTCCCATGT	80802		
Qy	1834	AGCTGGAAATACAGGACGACGCACATAGCCAGCTAATTTTGTATTTTACAGTAAACGG	1893		
Db	80803	AGCTGGATATACAGGACACACACAGCCACGCTAATTTTGTATTTTATAGTAGACAG	80866		
Qy	1894	GATTTACCAAGTGGCCAGCGGTGTCTGCACTCTGACCCGAAATATCCACCGGCT	1953		
Db	80863	GGTTTCACTGTGTGGCCAGCGGTGTCTCAACTCTCTGACCTCAATGATATCACCGGCT	80922		
Qy	1954	CCGCTCCCAAGTGTGGGATTTACAGGCGGTGAGCCACGCTGCGCGG	2002		
Db	80923	TAGCTTCAAAAGTCTGGGATTTACAGGACATGAGCCACTGCGCGG	80971		
RESULT 4					
US-09-608-285A-59					
Sequence 59, Application US/09608285A					
Patent No. 6335013					
GENERAL INFORMATION:					
APPLICANT: Ford, John					
APPLICANT: Mulero, Julio					
APPLICANT: Yeung, George					
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39- LIKE					
FILE REFERENCE: 28110/36570					
CURRENT APPLICATION NUMBER: US/09/608.285A					
CURRENT FILING DATE: 2000-06-30					
PRIOR APPLICATION NUMBER: 09/583.231					
PRIOR FILING DATE: 2000-05-26					
PRIOR APPLICATION NUMBER: 09/557.800					
PRIOR FILING DATE: 2000-04-25					
PRIOR APPLICATION NUMBER: 09/481.238					
PRIOR FILING DATE: 2000-01-11					
PRIOR APPLICATION NUMBER: 09/370.265					
PRIOR FILING DATE: 1999-08-09					
PRIOR APPLICATION NUMBER: PCT/US99/16180					
PRIOR FILING DATE: 1999-07-16					
PRIOR APPLICATION NUMBER: 09/350.836					
PRIOR FILING DATE: 1999-07-09					
PRIOR APPLICATION NUMBER: 09/273.447					
PRIOR FILING DATE: 1999-03-19					
PRIOR APPLICATION NUMBER: 09/244.444					
PRIOR FILING DATE: 1999-02-04					
PRIOR APPLICATION NUMBER: 09/122.449					

```

; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59
; LENGTH: 15977
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: CD39-L4/L6 Gene Sequence
; NAME/KEY: CDS
; LOCATION: (245)..(461)
; NAME/KEY: CDS
; LOCATION: (1454)..(1533)
; NAME/KEY: CDS
; LOCATION: (2734)..(2877)
; NAME/KEY: CDS
; LOCATION: (4364)..(4439)
; NAME/KEY: CDS
; LOCATION: (4679)..(4714)
; NAME/KEY: CDS
; LOCATION: (5326)..(5414)
; NAME/KEY: CDS
; LOCATION: (5723)..(5802)
; NAME/KEY: CDS
; LOCATION: (6751)..(6812)
; NAME/KEY: CDS
; LOCATION: (7758)..(7859)
; NAME/KEY: CDS
; LOCATION: (8712)..(8852)
; NAME/KEY: CDS
; LOCATION: (9831)..(9887)
; NAME/KEY: CDS
; LOCATION: (11613)..(11728)
; NAME/KEY: CDS
; LOCATION: (13146)..(13691)
; NAME/KEY: CDS
; LOCATION: (15702)..(15839)
; NAME/KEY: misc feature
; LOCATION: (114871)
; OTHER INFORMATION: n = a or c or g or t
; US-09-608-285A-59

Query Match          9.7%; Score 221.4; DB 4; Length 15977;
Best Local Similarity 87.0%; Pred. No. 1.7e-50;
Matches 255; Conservative 0; Mismatches 36; Indels 2; Gaps 1;

QY 1713 CTTTCTTTTGGAGAGGCGCTTTGCTTGTCAACCCAGCGTGAATGACAGTGGCGGGATCT 1772
DB 11237 CTTTCTTTTGGAGAGAGTCTTGTCTGTCTGTCAACCCAGCGTGAATGACAGTGGTGAATCT 11296
QY 1773 CAGCTACCGCAACTCTCACTCTGCGGTTCAAGTATTTTCTGCTCAAGCTTCCAG 1832
DB 11297 GCGCTACGTGAACCTCTGCTTCCAGGTTCAAGGATTTCTGCTCAAGCTTCCAG 11356
QY 1893 TAGCTGGGAATACA--GGCAAGCCCAACATGCCCACTAATTTTGTATTTTCAAGAGAA 1890
DB 11337 TAGCTGGGAATACAAGCGGTAACCAACAGCCCTCTAATTTTGTATTTTGAAGAGA 11416
QY 1891 GCGGATTTACCAACGTTGGCCAGGCTGTGCTGAACTCTGAACCCGAAGTATCCACCG 1950
DB 11417 TGGGGTTTACCAACGTTGGCCAGGCTGTGATCTGAATCTCTGACTTCAAGTATCCACCG 11476
QY 1951 CTTCCGCTTCCCAAGTGTGGGATTTACAGGCGTGAAGCCAGCCGCGGCC 2003
DB 11477 CTTCAAGCTTCCCAAGTGTGATTAAGGATTTACAGGCGTGAAGCCAGCCGCTGCGCC 11529

RESULT 5
US-09-608-285A-42
; Sequence 42, Application US/09608285A
; Patent No. 6335013
```

```

; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; APPLICANT: Yeung, George
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE
; FILE REFERENCE: 2810/36570
; CURRENT APPLICATION NUMBER: US/09/608,285A
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/583,231
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 09/557,800
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/481,238
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 09/370,265
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: 09/350,836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
; LENGTH: 14747
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (13641)
; OTHER INFORMATION: n = adenosine or guanine or cytosine or thymidine
; US-09-608-285A-42

Query Match          9.6%; Score 218.8; DB 4; Length 14747;
Best Local Similarity 86.6%; Pred. No. 8.3e-50;
Matches 253; Conservative 0; Mismatches 37; Indels 2; Gaps 1;

QY 1714 TTTTCTTTTGGAGAGGCGCTTTGCTTGTCAACCCAGCGTGAATGACAGTGGCGGATCTC 1773
DB 10008 TTTTCTTTTGGAGAGAGTCTTGTCTGTCTGTCAACCCAGCGTGAATGACAGTGGTGAATCTC 10067
QY 1774 AGCTACCGCAACCTCTGCTTCCAGGTTCAAGGATTTCTGCTCAAGCTTCCAG 1833
DB 10068 GCGTCACTGCAACCTCTGCTTCCAGGTTCAAGGATTTCTGCTCAAGCTTCCAG 10127
QY 1834 AGCTGGGAATACA--GGCAAGCCCAACATGCCCACTAATTTTGTATTTTCAAGAGAA 1891
DB 10128 AGCTGGGAATACAAGCGGTAACCAACAGCCCTCTAATTTTGTATTTTGAAGAGAT 10187
QY 1892 GGAATTTACCAACGTTGGCCAGGCTGTGCTGAACTCTGAACCCGAAGTATCCACCG 1951
DB 10188 GGGGTTTACCAACGTTGGCCAGGCTGTGATCTGAATCTCTGACTTCAAGTATCCACCG 10247
QY 1952 CTTCCGCTTCCCAAGTGTGGGATTTACAGGCGTGAAGCCAGCCGCGGCC 2003
DB 10248 CTCAGCTTCCCAAGTGTGATTAAGGATTTACAGGCGTGAAGCCAGCCGCTGCGCC 10299

RESULT 6
US-09-557-800C-42
; Sequence 42, Application US/09557800C
; Patent No. 6476211
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
```

APPLICANT: Young, George
TITLE OF INVENTION: Methods and Materials Relating to CD39-like
FILE REFERENCE: 28110/36457
CURRENT APPLICATION NUMBER: US/09/557,800C
CURRENT FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/481,238
PRIOR FILING DATE: 2000-01-11
PRIOR APPLICATION NUMBER: 09/370,265
PRIOR FILING DATE: 1999-08-09
PRIOR APPLICATION NUMBER: PCT/US99/16180
PRIOR FILING DATE: 1999-07-16
PRIOR APPLICATION NUMBER: 09/350836
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: 09/273447
PRIOR FILING DATE: 1999-03-19
PRIOR APPLICATION NUMBER: 09/122449
PRIOR FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: 09/244444
PRIOR FILING DATE: 1999-02-04
PRIOR APPLICATION NUMBER: 09/118,205
PRIOR FILING DATE: 1998-07-16
NUMBER OF SEQ ID NOS: 56
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 42
LENGTH: 14747
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (13641)
OTHER INFORMATION: n = adenosine or guanine or cytosine or thymidine
US-09-557-800C-42

Query Match 9.6%; Score 218.8; DB 4; Length 14747;
Best Local Similarity 86.6%; Pred. No. 8.3e-50;
Matches 253; Conservative 0; Mismatches 37; Indels 2; Gaps 1;

QY 1714 TTTTGGGAGAGGCGTTGCTTTGTCACCCAGCTGGAATGAGTGGCGGATCTC 1773
DB 10008 TTTTGGGAGAGGCGTTGCTTTGTCACCCAGCTGGAATGAGTGGCGGATCTC 10067
QY 1774 AGCTACCCGCACTCCGCTGCTGCAAGATTTTCCGCGCTCAGCTCCGCAAGT 1833
DB 10068 GGGTCTACGACCTCTGCTTCCAGTTCAAGGATTTCTCGCTCAGCTCCGCAAGT 10127
QY 1834 AGCTGGGAATACA--GGCAGCCGACCATGCGCAGCTAATTTTGTATTTTCACTAGAAAC 1891
DB 10128 AGCTGGGATTAACAGCGGTACACCAACGCGGTCTAATTTTGTATTTTGTAGAGAT 10187
QY 1892 GGGATTTACCAAGTGGCGGAGGTGTCTGAACTCTGACCGCAAGTATCCACCGGC 1951
DB 10188 GGGGTTTACCAAGTGGCGGAGGTGTCTGAACTCTGACCGCAAGTATCCACCGGC 10247
QY 1952 CTCGGCTCCCAAGTGTGGGATTAAGGCGTGAAGCCAGCGCCGCC 2003
DB 10248 CTCAGCTCCCAAGTGTGGGATTAAGGCGTGAAGCCAGCGCCGCC 10299

RESULT 7
US-09-800-960-3/c
Sequence 3, Application US/09800960
Patent No. 6387677
GENERAL INFORMATION:
APPLICANT: YE, Jane et al.
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
FILE REFERENCE: CL001158
CURRENT APPLICATION NUMBER: US/09/800,960
CURRENT FILING DATE: 2001-03-08
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FaestSeq for Windows Version 4.0

SEQ ID NO 3
LENGTH: 62804
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1) (62804)
OTHER INFORMATION: n = A,T,C or G
US-09-800-960-3

Query Match 9.6%; Score 218.6; DB 4; Length 62804;
Best Local Similarity 86.1%; Pred. No. 1.9e-49;
Matches 254; Conservative 0; Mismatches 39; Indels 2; Gaps 1;

QY 1714 TTTTGGGAGAGGCGTTGCTTTGTCACCCAGCTGGAATGAGTGGCGGATCTC 1773
DB 13733 TTTCTTTGAGACGAGGAGTCTGTTGTCAGGCGTGGGAGTGAATGATCTC 13674
QY 1774 AGCTACCCGCACTCCGCTGCTGCAAGTATTTTCTGCTCAGCTCCGCAAGT 1833
DB 13673 AGCTACCTCAACCTTGGCTCTGCTGCAAGTATTTTCTGCTCAGCTCCGCAAGT 13614
QY 1834 AGCTGGGAATACAAG--CAGCCACCATGCGCAGCTAATTTTGTATTTTCACTAGAAAC 1891
DB 13613 AGCTGGGATTAACAGGACACACACAGCGCTGCTAATTTTGTATTTTGTAGAGAT 13554
QY 1892 GGGATTTACCAAGTGGCGGAGGTGTCTGAACTCTGACCGCAAGTATCCACCGGC 1951
DB 13553 GGGGTTTACCAAGTGGCGGAGGTGTCTGAACTCTGACCGCAAGTATCCACCGGC 13494
QY 1952 CTCGGCTCCCAAGTGTGGGATTAAGGCGTGAAGCCAGCGCCGCC 2006
DB 13493 CTCGGCTCCCAAGTGTGGGATTAAGGCGTGAAGCCAGCGCCGCC 13439

RESULT 8
US-08-579-445-26/c

Sequence 26, Application US/08579445
Patent No. 6566053
GENERAL INFORMATION:
APPLICANT: Perucho, Manuel
APPLICANT: Perinado, Miguel A.
APPLICANT: Ionov, Yuri
APPLICANT: Malkhosyan, Sergei
TITLE OF INVENTION: Identification of Neoplasms by Detection
NUMBER OF SEQUENCES: 27
TITLE OF INVENTION: of Genetic Deletions
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/579,445
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/152,484
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Kirkpatrick, Anita M.
REGISTRATION NUMBER: 32,617
REFERENCE/DOCKET NUMBER: STRATAG. 009A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 283 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-579-445-26

Query Match
Best Local Similarity 9.6%; Score 218.4; DB 4; Length 283;
Pred. No. 1.5e-50;
Matches 249; Conservative 0; Mismatches 31; Indels 2; Gaps 1;

1723 GAGAGGCGCTCTGTTGTCACCCGCTGGAATGAGTGGGGATCTCAGCTACCG 1782
|||||
282 GAGAGGAGCTCTGCTCTCTCGCCGAGTGAGTGCAGTGGCGCATCTCGGCTCAGT 223
1783 CAACCTCCACCTCTGAGTTCAAGTATTTCTGCTCAGCCTCCCAAGTAGTGGAA 1842
|||||
222 CAACCTCCACCTCCGAGTTCAAGCATTTCTGCTCAGCCTCCCGAGTAGCTGGAT 163
1843 TACAG--GCAGCCACCATGCCAGCTAATTTTGTATTTTCAGTAGAAACGGGATTTCA 1900
|||||
162 TACAGGCGCGCCACACGCGGCTAATTTTGTATTTTATAGTAGACGGGGTTTCA 103
1901 CCAGCTGGCCAGGCTGCTGTAATCTCTGACCGGAATGATCCACCCGCTCGGCTTC 1960
|||||
102 CATGTGTGCGCAGGCTGTTGAACTCTGACTGAGTGTATCCACCATCTCGGCTTC 43
1961 CCAAGTGTGGATTACAGGCGGTGAGCCAGCCGTCGCGGCG 2002
42 CCAAGTGTGGATTACAGGCTGAGTGTAGCCACCATCCAGCCAGCC 1

Db

RESULT 9
US-08-133-629-8/c
Sequence 8, Application US/08133629
Patent No. 5597694

GENERAL INFORMATION:
APPLICANT: Munroe, David J.
APPLICANT: Housman, David E.
TITLE OF INVENTION: AMPLIFICATION OF NUCLEIC ACIDS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 02210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk.
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/133,629
FILING DATE: 07-OCT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Greer, Helen
REGISTRATION NUMBER: 36,816
REFERENCE/DOCKET NUMBER: M0828/7001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX: 92-1742 EZEKIEL
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
US-08-133-629-8

Query Match
Best Local Similarity 9.6%; Score 218; DB 1; Length 282;
Pred. No. 1.9e-50;
Matches 245; Conservative 5; Mismatches 30; Indels 2; Gaps 1;

1723 GAGAGGCGCTCTGTTGTCACCCAGCTGGAATGAGTGGGGATCTCAGCTACCG 1782
|||||
282 GAGAGGAGCTCTGCTCTCTCGCCGAGTGAGTGCAGTGGCGCATCTCGGCTCAGT 223
1783 CAACCTCCACCTCTGAGTTCAAGTATTTCTGCTCAGCCTCCCAAGTAGTGGAA 1842
|||||
222 CAACCTCCACCTCTGAGTTCAAGCATTTCTGCTCAGCCTCCCGAGTAGCTGGAT 163
1843 TACAG--GCAGCCACCATGCCAGCTAATTTTGTATTTTTCAGTAGAAACGGGATTTCA 1900
|||||
162 TACAGGCGCGCCACACGCGGCTAATTTTGTATTTTATAGTAGACGGGGTTTCA 103
1901 CCAGCTGGCCAGGCTGCTGTAATCTCTGACCGGAATGATCCACCCGCTCGGCTTC 1960
|||||
102 CATGTGTGCGCAGGCTGTTGAACTCTGACTGAGTGTATCCACCATCTCGGCTTC 43
1961 CCAAGTGTGGATTACAGGCGGTGAGCCAGCCGTCGCGGCG 2002
42 CCAAGTGTGGATTACAGGCTGAGTGTAGCCACCATCCAGCCAGCC 1

Db

RESULT 10
US-08-520-373D-4
Sequence 4, Application US/08520373D
Patent No. 6451753

GENERAL INFORMATION:
APPLICANT: Tombran-Tink, Joyce
APPLICANT: Steele, Flintan R
APPLICANT: Chader, Gerald J
APPLICANT: Becerra, Sofia P
APPLICANT: Johnson, Lincoln V
APPLICANT: Rodriguez, Ignacio R
TITLE OF INVENTION: RETINAL PIGMENTED EPITHELIUM DERIVED NEUROTROPIC FACTOR
FILE REFERENCE: 2026-4203US1
CURRENT APPLICATION NUMBER: US/08/520,373D
CURRENT FILING DATE: 1995-08-29
PRIOR APPLICATION NUMBER: 08/377,710
PRIOR FILING DATE: 1995-01-25
PRIOR APPLICATION NUMBER: 08/279,979
PRIOR FILING DATE: 1994-07-25
PRIOR APPLICATION NUMBER: 07/894,215
PRIOR FILING DATE: 1992-06-04
PRIOR APPLICATION NUMBER: 07/952,796
PRIOR FILING DATE: 1992-09-24
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4

LENGTH: 14581
TYPE: DNA
ORGANISM: HUMAN
FEATURE:
OTHER INFORMATION: mRNA: 6683; EXON: 6683-6790; EXON 11584-11675;
OTHER INFORMATION: EXON: 14539-14581; INTRON: 6791-11583; INTRON:
OTHER INFORMATION: 11676-14538; CDS: 11584-11675; 14539-14580
US-08-520-373D-4

Query Match
Best Local Similarity 9.6%; Score 217.8; DB 4; Length 14581;
Pred. No. 1.6e-49;
Matches 252; Conservative 0; Mismatches 37; Indels 2; Gaps 1;

1714 TTTTGGAGAGAGGCTCTGCTTGTGTCACCCAGCTGGAATGAGTGGGGATCTC 1773
|||||
9530 TTTTGGAGATGAGATCTGCTCTGTGCGCCAGCTGAGTGCAGTGGCGCATCTTC 9589
1774 AGCTACCGCAACCTCCCTGCTTCAAGTATTTTCTGCTCAGCTCCCAAGT 1833

Db 9590 AGTCACTGCAACCTCCTCTGATTCATTAATGATTCCTCCGCTTACCTCCCAAGT 9649
QY 1834 AGTGGGAATACAGGCA--CGCCACCATGCCAGCTAATTTTGTATTTTCAGTAGAAG 1891
Db 9650 AGTGGGATTAACAGGATGCGCCACACCCGGCTAATTTTGTATTTTGTAGAGAT 9709
QY 1892 GGGATTTCACCAAGTGGGCGAGGCTGTCTCGAATCTCGGCAAGTATCCACCCGC 1951
Db 9710 GGGGTTTCTCAGTGTGGCCAGGATGCTCAAACTCTGACTTCAGGTATCTACCCGC 9769
QY 1952 CTCGGCTCCCAAGTCTGGGATTAAGGCGTAGCCACCGTCCCGGCC 2002
Db 9770 CTCGGCTCTCAAGTGTGGATTAAGGTTTGAAGCACTGCGCTCGCC 9820

RESULT 11
US-08-367-841A-43

; Sequence 43, Application US/08367841A
; Patent No. 6319687
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Rodriguez,
; APPLICANT: Ignacio R.; Mazuruk, Krzysztof;
; APPLICANT: Tombrax-Tink, Joyce
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
; TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/367,841A
; FILING DATE: 30-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,963
; FILING DATE: 07-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36434
; REFERENCE/DOCKET NUMBER: 20264126US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22481 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: P1-147
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: full length genomic
; OTHER INFORMATION: sequence for PEDF plus flanking sequences.
US-08-367-841A-43

Query Match 9.6%; Score 217.8; DB 4; Length 22481;

Best Local Similarity 86.6%; Pred. No. 1,9e-49;
Matches 222; Conservative 0; Mismatches 37; Indels 2; Gaps 1;

QY 1714 TTTTGTGGAGAGGCGTCTTGTCTTGTACCCACGCTGAATGAGTGGGGATCTC 1773
Db 9519 TTTTGTGGAGATGAGTCTTGTCTTGTGTGCGCAGCTGAGTGAAGTGGCGATCTC 9578
QY 1774 AGTCAACCGCAACTCCACCTCTGAGTATTTTCTGCTAGCTTCCGCTCCCAAGT 1833
Db 9579 AGTCACTCAACCTCCTCCTCTGATTAAGTATCTCTGCTTAGCTCCCAAGT 9638
QY 1834 AGTGGGAATACAGGCA--CGCCACCATGCCAGCTAATTTTGTATTTTCAGTAGAAG 1891
Db 9639 AGTGGGATTAACAGGATGCGCCACACCCGGCTAATTTTGTATTTTGTAGAGAT 9698
QY 1892 GGGATTTCACCAAGTGGGCGAGGCTGTCTCGAATCTCTGACCGCAAGTATCCACCCGC 1951
Db 9699 GGGGTTTCTCAGTGTGGCCAGGATGCTCAAACTCTGACTTCAGGTATCTACCCGC 9758
QY 1952 CTCGGCTCCCAAGTCTGGGATTAAGGCGTAGCCACCGTCCCGGCC 2002
Db 9759 CTCGGCTCTCAAGTGTGGATTAAGGTTTGAAGCACTGCGCTCGCC 9809

RESULT 12

PCT-US95-07201-43
; Sequence 43, Application PC/TUS9507201

; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Becerra, Sofia
; APPLICANT: Patricia; Schwartz, Joan P.;
; APPLICANT: Taniwaki, Takayuki
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
; TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07201
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/367,841
; FILING DATE: 30-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,963
; FILING DATE: 07-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36434
; REFERENCE/DOCKET NUMBER: 20264126PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22481 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown

MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Pl-147
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: Full length genomic
OTHER INFORMATION: Sequence for PEDF plus flanking sequences.
PCT-US95-07201-43

Query Match
Best Local Similarity 9.6%; Score 217.8; DB 5; Length 22484;
Matches 252; Conservative 0; Mismatches 37; Indels 2; Gaps 1;

QY 1714 TTTTGTGGAGAGCGCTGCTTGTGTCACCCAGCTGAATGAGTGGGAGATCTC 1773
DB 9519 TTTTGTGGAGATGAGTCTGCTGCTGTCGCCACCTGAGTGAAGGCGCATCTC 9578
QY 1774 AGCTACCCGCAACCTCCACCTCTGGGTTCAAGTATTTTCTGCTCAGCTCCCAAGT 1833
DB 9579 AGCTACCTGCAACCTCCACCTCTGGGTTCAAGTATTTTCTGCTCAGCTCCCAAGT 9638
QY 1834 AGCTGGGATATACAGCA--CGCACCATGCCAGCTAATTTTGTATTTTTCAGTAAAC 1891
DB 9639 AGCTGGGATATACAGCAAGCATGCGCACCAACCGGCTAATTTTGTATTTTGTAGTAAAT 9698
QY 1892 GGGATTTCACACGTTGGCAGGCTGTCTCGAATCTGACCGCAAGTATCCACCGCC 1951
DB 9699 GGGGTTTCTCCATGTTGGCAGGATGTCTCAAACTCCGACTCAAGTATCTACCCGC 9758
QY 1952 CTCGCCCTCCCAAGTGTGCTGGATTAACAGGCGTGAACCGTCCCGGCC 2002
DB 9759 CTCGCCCTCTCAAGTGTGCTGGATTAACAGGCGTGAACCGTCCCGGCC 9809

RESULT 13

US-09-875-223-2
Sequence 2, Application US/09875223
Patent No. 6391850
GENERAL INFORMATION:
APPLICANT: No. 6391850, Western University
APPLICANT: David Dawson
APPLICANT: Paul Gillette
TITLE OF INVENTION: Methods and Compositions for Inhibiting Angiogenesis
FILE REFERENCE: 0290-2303
CURRENT APPLICATION NUMBER: US/09/875,223
CURRENT FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: US 09/122,079
PRIOR FILING DATE: 1998-07-23
PRIOR APPLICATION NUMBER: PCT/US98/15228
PRIOR FILING DATE: 1998-07-23
PRIOR APPLICATION NUMBER: US 08/899,304
PRIOR FILING DATE: 1997-07-23
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 2
LENGTH: 22484
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Unsure
LOCATION: 1...22484
OTHER INFORMATION: "n" means either a, c, t, or g
US-09-875-223-2

Query Match
Best Local Similarity 9.6%; Score 217.8; DB 4; Length 22484;
Matches 252; Conservative 0; Mismatches 37; Indels 2; Gaps 1;

QY 1714 TTTTGTGGAGAGCGCTGCTTGTGTCACCCAGCTGAATGAGTGGGAGATCTC 1773
DB 9519 TTTTGTGGAGATGAGTCTGCTGCTGTCGCCACCTGAGTGAAGGCGCATCTC 9578

QY 1774 AGCTACCCGCAACCTCCACCTCTGGGTTCAAGTATTTTCTGCTCAGCTCCCAAGT 1833
DB 9579 AGCTACCTGCAACCTCCACCTCTGGGTTCAAGTATTTTCTGCTCAGCTCCCAAGT 9638
QY 1834 AGCTGGGATATACAGCA--CGCACCATGCCAGCTAATTTTGTATTTTTCAGTAAAC 1891
DB 9639 AGCTGGGATATACAGCAAGCATGCGCACCAACCGGCTAATTTTGTATTTTGTAGTAAAT 9698
QY 1892 GGGATTTCACACGTTGGCAGGCTGTCTCGAATCTGACCGCAAGTATCCACCGCC 1951
DB 9699 GGGGTTTCTCCATGTTGGCAGGATGTCTCAAACTCCGACTCAAGTATCTACCGCC 9758
QY 1952 CTCGCCCTCCCAAGTGTGCTGGATTAACAGGCGTGAACCGTCCCGGCC 2002
DB 9759 CTCGCCCTCTCAAGTGTGCTGGATTAACAGGCGTGAACCGTCCCGGCC 9809

RESULT 14

US-09-820-002-3
Sequence 3, Application US/09820002
Patent No. 6482630
GENERAL INFORMATION:
APPLICANT: Gan, Weiniu
APPLICANT: Ye, Jane
APPLICANT: DiFrancesco, Valentina
APPLICANT: Beasley, Ellen
TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
FILE REFERENCE: CLO01194
CURRENT APPLICATION NUMBER: US/09/820,002
CURRENT FILING DATE: 2001-03-29
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 21784
TYPE: DNA
ORGANISM: HUMAN
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(21784)
OTHER INFORMATION: n = A,T,C or G
US-09-820-002-3

Query Match
Best Local Similarity 9.5%; Score 215.6; DB 4; Length 21784;
Matches 263; Conservative 0; Mismatches 59; Indels 2; Gaps 1;

QY 1714 TTTTGTGGAGAGCGCTGCTTGTGTCACCCAGCTGAATGAGTGGGAGATCTC 1773
DB 18045 TTTTGTGGAGAGAGATCTTCTGCTGTCGCCAGCTGAGTCAATGACATATCTC 18104
QY 1774 AGCTACCCGCAACCTCCACCTCTGGGTTCAAGTATTTTCTGCTCAGCTCCCAAGT 1833
DB 18105 GGGTCACTGCAACCTCCACCTCTGGGTTCAAGTATTTTCTGCTCAGCTCCCAAGT 18164
QY 1834 AGCTGGGATATACAG--GCAAGCCACCATGCCAGCTAATTTTGTATTTTTCAGTAAAC 1891
DB 18165 AGCTGGGATATACAGTGTGCTGCAACCATGCCAGCTAATTTTGTATTTTGTAGTAAAT 18224
QY 1892 GGGATTTCACACGTTGGCAGGCTGTCTCGAATCTGACCGCAAGTATCCACCGCC 1951
DB 18225 GAGGTTTGGCATGTTGGCAGGCTGTCTCAAACTCTGACATCAAGTATCTCTCTGC 18284
QY 1952 CTCGCCCTCCCAAGTGTGCTGGATTAACAGGCGTGAACCGTCCCGGCC 2011
DB 18285 CTCGCCCTCCCAAGTGTGCTGGATTAACAGGCGTGAACCGTCAATGATGTGTGT 18344
QY 2012 AACTCTTGTGGAGAGAGAGAGG 2035
DB 18345 TTTTGTGGAGAGAGAGAGAGG 18368

```

RESULT 15
US-09-810-347-3
; Sequence 3, Application US/09810347
; Patent No. 6461847
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001169
; CURRENT APPLICATION NUMBER: US/09/810,347
; CURRENT FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 42571
; TYPE: DNA
; ORGANISM: Human
US-09-810-347-3

```

```

Query Match          9.5%; Score 215.6; DB 4; Length 42571;
Best Local Similarity 86.0%; Pred. No. 1.1e-48;
Matches 251; Conservative 0; Mismatches 39; Indels 2; Gaps 1;

QY 1717 TTTTGGAGAGGCGCTTGTCTTCACCCAGCTGGAATGCAAGTGGCGGATCTCAGC 1776
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 36537 TTTTGGAGATAGAGTCTGCTGTACCCAGGCTGAGTGCAGTGGCGCAATCTCGGC 36596

QY 1777 TCACCGCAACCTCCACCTCCTGGGTTCAAGTATTTTCTGCTCAGCCTCCCAAGTAC 1836
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 36597 TCACCTGCAACCTCCACCTCCTGTGTTCAAGATTTCTGCTCAGCCTCTGAGTAGC 36656

QY 1837 TGGGAATACG--GCAAGCCACCATGCGCACTAATTTTGTATTTTTCAGTAGAAAGGG 1894
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 36657 TGGGATTAACAGGCGCCGCGCACCATGCGCACTAATTTTGTATTTTTCAGTAGAAAGGG 36716

QY 1895 ATTTCACAGTGGCGAGGCTGTCTGAACTCTGACCGCAAGTATCCACCGCTC 1954
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 36717 GTTTCCTCATGTGGCGAGGCTGTCTGAACTCTGAACTCTGAACTCTGAACTCTGAACTCTG 36776

QY 1955 CGCTCCCAAGTGTGAGTATACAGGCGTACGCAACCGTCCCGGCCAA 2006
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 36777 GGCCTTCCAAATGCTGGGATTAACAGGCGTACGCAACCATGCTGCCCAA 36828

```

Search completed: November 22, 2003, 07:54:51
 Job time : 138 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using ew model

Run on: November 22, 2003, 10:43:46 ; Search time 704 Seconds
(without alignments)
10552.029 Million cell updates/sec

Title: US-09-938-803-26

Perfect score: 2273

Sequence: 1 99999tcgaagcatcgtt.....tttccaaaaa 2273

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 2169961 seqs, 1634102185 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2273	100.0	2273	9	US-09-938-803-26
2	1842	81.0	1842	10	US-09-880-107-3415
3	1842	81.0	1842	10	US-09-954-531-1368
4	708	31.1	7461	9	US-09-764-860-1135
5	708	31.1	7461	10	US-09-764-904-91
6	708	31.1	7461	14	US-10-091-548-91
7	708	31.1	7461	14	US-10-074-095-1135
8	708	31.1	32174	9	US-09-908-711-158
9	708	31.1	32174	9	US-09-764-860-1134
10	708	31.1	32174	10	US-09-860-877-2645
11	708	31.1	32174	10	US-09-860-670-232
12	708	31.1	32174	10	US-09-764-904-90
13	708	31.1	32174	11	US-09-764-891-6480
14	708	31.1	32174	11	US-09-764-891-10135
15	708	31.1	32174	11	US-09-764-891-10179
16	708	31.1	32174	14	US-10-091-548-90

C 17	708	31.1	32174	14	US-10-074-095-1134	Sequence 1134, Ap
18	428	18.8	459	11	US-09-918-995-22426	Sequence 22426, A
19	392	17.2	475	11	US-09-918-995-26257	Sequence 26257, A
20	355	15.6	363	11	US-09-918-995-30086	Sequence 30086, A
21	317	13.9	490	11	US-09-918-995-11899	Sequence 11899, A
22	254	11.2	990	10	US-09-964-824A-271	Sequence 271, App
23	254	11.2	990	10	US-09-954-531-973	Sequence 973, App
24	254	11.2	990	10	US-09-954-531-1369	Sequence 1369, App
25	181	8.0	226	12	US-10-029-386-16123	Sequence 16123, A
26	181	8.0	505	12	US-10-029-386-2423	Sequence 2423, A
27	167	7.3	173	12	US-10-029-386-16861	Sequence 16861, A
28	167	7.3	599	12	US-10-029-386-3161	Sequence 3161, Ap
29	163	7.2	478	9	US-09-864-761-87	Sequence 87, App1
30	153	6.7	758	12	US-10-027-632-8362	Sequence 8362, Ap
31	153	6.7	758	13	US-10-027-632-8362	Sequence 8362, Ap
32	153	6.7	2006	12	US-10-027-632-89441	Sequence 99441, A
33	153	6.7	2006	13	US-10-027-632-99441	Sequence 99441, A
34	135	5.9	489	11	US-09-918-995-1933	Sequence 1933, Ap
35	128	5.6	177	9	US-09-864-761-16925	Sequence 16925, A
36	112	4.9	217	10	US-09-796-692-7696	Sequence 7696, Ap
37	112	4.9	217	14	US-10-040-862-7696	Sequence 7696, Ap
38	73	3.2	280	10	US-09-880-107-1992	Sequence 1992, Ap
39	59	2.6	74962	14	US-10-274-974-3	Sequence 3, App1
40	55	2.4	923	12	US-10-027-632-10590	Sequence 10590, A
41	55	2.4	923	13	US-10-027-632-10590	Sequence 10590, A
42	53	2.3	10901	11	US-09-764-891-5924	Sequence 5924, Ap
43	52	2.3	819	12	US-10-027-632-140838	Sequence 140838, A
44	52	2.3	819	13	US-10-027-632-140838	Sequence 140838, A
45	52	2.3	76798	10	US-09-880-107-3949	Sequence 3949, Ap

ALIGNMENTS

RESULT 1
US-09-938-803-26
; Sequence 26, Application US/09938803
; Patent No. US20020076762A1
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Tang, Y. Tom
; APPLICANT: Lal, Preeti
; APPLICANT: Reddy, Roopa
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Yang, Junming
; APPLICANT: Azimzal, Yalda
; TITLE OF INVENTION: FULL-LENGTH EXPRESSED GENETIC MARKERS
; FILE REFERENCE: PF-0695 US
; CURRENT APPLICATION NUMBER: US/09/938,803
; CURRENT FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/311,894
; PRIOR FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 26
; LENGTH: 2273
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte Clone 1928920
US-09-938-803-26

Query Match 100.0%; Score 2273; DB 9; Length 2273;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGTGAAGCATGCTTTGGCCGCACTGGGGGCGGCGGAGTGGGGGGTCCCTGCG 60
DB 1 GGGGGTGAAGCATGCTTTGGCCGCACTGGGGGCGGCGGAGTGGGGGGTCCCTGCG 60
QY 61 GGGCTCCGAGATTAGATGGCTTCAGCGGAGGGGAGCAGAGGAGACTGTGTGTCGCG 120

Db 61 GGGCTCCGGAGTTAAGATGGCGTCTCAGCGAGGGGGAAGAAGGAGCTGTGTGGCCG 120
Qy 121 TGGCGGGGGTTCTGCACTCGGGTTTCCAGAGCTGAGCCTTAAACAAGTTGGCACTGCC 180
Db 121 TGGCGGGGGTTCTGCACTCGGGTTTCCAGAGCTGAGCCTTAAACAAGTTGGCACTGCC 180
Qy 181 TGGCGGGGGTGAAGAACGGGGCTGGGCTGATCATCTCTCTGGGTTACCCCTTTG 240
Db 181 TGGCGGGGGTGAAGAACGGGGCTGGGCTGATCATCTCTCTGGGTTACCCCTTTG 240
Qy 241 CTCTGTTTATCGGCATTACCTTTTTCAGAGAGACCTACCTACCTACCTCTTCCATA 300
Db 241 CTCTGTTTATCGGCATTACCTTTTTCAGAGAGACCTACCTACCTACCTCTTCCATA 300
Qy 301 CCTTTACAGGCTCTCAATTGCTTTATTTTAACTTTGAAACAGCTTACCACTCCCTGC 360
Db 301 CCTTTACAGGCTCTCAATTGCTTTATTTTAACTTTGAAACAGCTTACCACTCCCTGC 360
Qy 361 TGTGTATGTGCTTCACTTCTCATCTTGAAGTGGGCGGACCATCATGCGGCTC 420
Db 361 TGTGTATGTGCTTCACTTCTCATCTTGAAGTGGGCGGACCATCATGCGGCTC 420
Qy 421 TCACTACCTTTTCTTCCAGATGGCTTACCTTCTGCTGATATCTATTACACTGCCACG 480
Db 421 TCACTACCTTTTCTTCCAGATGGCTTACCTTCTGCTGATATCTATTACACTGCCACG 480
Qy 481 GCACATACGATATCAAGTGAACAATGCAATGCTGTCTGACTTTGAACTGATGGTT 540
Db 481 GCACATACGATATCAAGTGAACAATGCAATGCTGTCTGACTTTGAACTGATGGTT 540
Qy 541 TGGCTGTTGACTCTTTGACGAGGAGGAAGATCAGAAATTCCTGTCTGAGCAACAGA 600
Db 541 TGGCTGTTGACTCTTTGACGAGGAGGAAGATCAGAAATTCCTGTCTGAGCAACAGA 600
Qy 601 AATATGCAATACGTGTGTCTCTCCCTGCTGGAAGTGTGTGTCTCTTCTTCTATG 660
Db 601 AATATGCAATACGTGTGTCTCTCCCTGCTGGAAGTGTGTGTCTCTTCTTCTATG 660
Qy 661 GGGCTCTTGTGTAGGGGCCCCAGTCTCATGTAATCTACATGAAGCTGTGAGGAG 720
Db 661 GGGCTCTTGTGTAGGGGCCCCAGTCTCATGTAATCTACATGAAGCTGTGAGGAG 720
Qy 721 AGCTGATTGACATACAGAGAAAGATACCAACAGCATCATTTCTGCTCAAGGCTGTA 780
Db 721 AGCTGATTGACATACAGAGAAAGATACCAACAGCATCATTTCTGCTCAAGGCTGTA 780
Qy 781 GTCTGGGCTTTTCTACCTAGTGGCTACACACTGCTCAGCCCCACATCAGAGAACT 840
Db 781 GTCTGGGCTTTTCTACCTAGTGGCTACACACTGCTCAGCCCCACATCAGAGAACT 840
Qy 841 ATCTCCCTACGAAAGATATGACAAACACCCCTTCTGCTCCGCTGATGATGCTGA 900
Db 841 ATCTCCCTACGAAAGATATGACAAACACCCCTTCTGCTCCGCTGATGATGCTGA 900
Qy 901 TCTGGGCAAGTTGTCTGTACAAATATGTCACTTGTGCTGTGACAGAGAGATAT 960
Db 901 TCTGGGCAAGTTGTCTGTACAAATATGTCACTTGTGCTGTGACAGAGAGATAT 960
Qy 961 GCATTTTGAAGGGCTGTGGCTTCAATGCTTTGAAAGAAAGGGCAAGGCAAGTGGATG 1020
Db 961 GCATTTTGAAGGGCTGTGGCTTCAATGCTTTGAAAGAAAGGGCAAGGCAAGTGGATG 1020
Qy 1021 CCTGTGCAACATGAAGGTGTGCTTTTGAAGAAACCCCGCTTACCTGAGCACTTGG 1080
Db 1021 CCTGTGCAACATGAAGGTGTGCTTTTGAAGAAACCCCGCTTACCTGAGCACTTGG 1080
Qy 1081 CCTCATTCACATCAACCAAGCGCTGGGTGCGGCTGATCTTTCAAGAGATCAAGT 1140
Db 1081 CCTCATTCACATCAACCAAGCGCTGGGTGCGGCTGATCTTTCAAGAGATCAAGT 1140
Qy 1141 TCTTGGAAATTAAGAACTCTCTCAGGGTCTCTGTTGTATTTCTGAGCTCTGGACG 1200
Db 1141 TCTTGGAAATTAAGAACTCTCTCAGGGTCTCTGTTGTATTTCTGAGCTCTGGACG 1200

Qy 1201 GCTTGACTCAGATATCTGTGCTTCTTCCAGATGGAATTCCTCATTTGTTATGTGAAA 1260
Db 1201 GCTTGACTCAGATATCTGTGCTTCTTCCAGATGGAATTCCTCATTTGTTATGTGAAA 1260
Qy 1261 GACAGCTGCGAGGCTCATTTCAAGAGAGGCCCACTGTAGCAAGCTGGCGGCTATAGT 1320
Db 1261 GACAGCTGCGAGGCTCATTTCAAGAGAGGCCCACTGTAGCAAGCTGGCGGCTATAGT 1320
Qy 1321 TCTTCCAGGCTTCTTCACTATTTGTGCAACAGCATCACTGCTTTCATGGGTTACT 1380
Db 1321 TCTTCCAGGCTTCTTCACTATTTGTGCAACAGCATCACTGCTTTCATGGGTTACT 1380
Qy 1381 CCATGATGCTTCTGCTTCTTCACTGAGGAGCAAAAGGCTTAAAGGTATTAATCCATCT 1440
Db 1381 CCATGATGCTTCTGCTTCTTCACTGAGGAGCAAAAGGCTTAAAGGTATTAATCCATCT 1440
Qy 1441 ATTTCTTGGCCATCTTCTTCTGAGCTACTATTCATATTTGCTTATTCACAAAG 1500
Db 1441 ATTTCTTGGCCATCTTCTTCTGAGCTACTATTCATATTTGCTTATTCACAAAG 1500
Qy 1501 CAATGCTGCCAAGAAAGAAAGTTAAGAAAGTGAATATTCATTTCTTGTGGCTT 1560
Db 1501 CAATGCTGCCAAGAAAGAAAGTTAAGAAAGTGAATATTCATTTCTTGTGGCTT 1560
Qy 1561 GTGCGGAGCTGTGTCAGAAATCTACTGCTCTTTCACAGACTCTTTGCCCAAG 1620
Db 1561 GTGCGGAGCTGTGTCAGAAATCTACTGCTCTTTCACAGACTCTTTGCCCAAG 1620
Qy 1621 CAGAGAAATGAAAGAGCAGAGAGGTGAAGATGATGATTCAGCTGTGCTGTGCTGC 1680
Db 1621 CAGAGAAATGAAAGAGCAGAGAGGTGAAGATGATGATTCAGCTGTGCTGTGCTGC 1680
Qy 1681 AGCAAGCTCTTCAATTTGGGGCCAAAGGGGAACTTTTGTGGAAGAGCTTCTCTT 1740
Db 1681 AGCAAGCTCTTCAATTTGGGGCCAAAGGGGAACTTTTGTGGAAGAGCTTCTCTT 1740
Qy 1741 GTTACCCAGCTGTGAATGAGTGGCGGATCTCAGCTCACCCGCACTCTCCTGG 1800
Db 1741 GTTACCCAGCTGTGAATGAGTGGCGGATCTCAGCTCACCCGCACTCTCCTGG 1800
Qy 1801 TTCAATGATTTTCTGCTCAGCTCAGCTCCCAAGATAGCTGGGAATACAGGACGCCATG 1860
Db 1801 TTCAATGATTTTCTGCTCAGCTCAGCTCCCAAGATAGCTGGGAATACAGGACGCCATG 1860
Qy 1861 CCCAGCTAATTTTGTATTTTCAAGTAAACGGGATTTCAACAAGCTTGGCCAGCTGTCT 1920
Db 1861 CCCAGCTAATTTTGTATTTTCAAGTAAACGGGATTTCAACAAGCTTGGCCAGCTGTCT 1920
Qy 1921 TCGAATCTCGTGAACCGGAATGCCACCGGCTCCGCTCCCAAGATGCTGGGATTAAG 1980
Db 1921 TCGAATCTCGTGAACCGGAATGCCACCGGCTCCGCTCCCAAGATGCTGGGATTAAG 1980
Qy 1981 GCGTGAAGCAACCGTGCAGGCCCCAAGGGGAACTCTTGTGGAGAGAGAGGGGCTCA 2040
Db 1981 GCGTGAAGCAACCGTGCAGGCCCCAAGGGGAACTCTTGTGGAGAGAGAGGGGCTCA 2040
Qy 2041 CATCTCCCTCTGATTTCCCATGCAACTTGTCTTATCTCTCCCATCTAGCAGGAATC 2100
Db 2041 CATCTCCCTCTGATTTCCCATGCAACTTGTCTTATCTCTCCCATCTAGCAGGAATC 2100
Qy 2101 TATTTGTTTTTCTTGTGCAATTTTACTATGATGTGATGTGCGCTTACCAACACCCC 2160
Db 2101 TATTTGTTTTTCTTGTGCAATTTTACTATGATGTGATGTGCGCTTACCAACACCCC 2160
Qy 2161 CCCATGGGGGGGTGAGAGGGGTGCAAGGCTGTGCTGTCACTTTTCTACTTTGAA 2220
Db 2161 CCCATGGGGGGGTGAGAGGGGTGCAAGGCTGTGCTGTCACTTTTCTACTTTGAA 2220
Qy 2221 CTGTATTAGATTAATCACTTCTGTTGTTCAGTTTTCATTTTCAAAAAA 2273
Db 2221 CTGTATTAGATTAATCACTTCTGTTGTTCAGTTTTCATTTTCAAAAAA 2273

RESULT 2
US-09-880-107-3415
Sequence 3415, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 3415
LENGTH: 1842
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 U72515
US-09-880-107-3415

Query Match 81.0%; Score 1842; DB 10; Length 1842;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 279 TACCTCATCCACCTCTTCCATCCTTACAGGCTCTCAATGCTATTTTAACTTTGA 338
Db 1 TACCTCATCCACCTCTTCCATCCTTACAGGCTCTCAATGCTATTTTAACTTTGA 60
Qy 339 AACCAAGCTTACCACTCCCTGTGTGTATTTGCTTCACTTCTCAATCCTTGAAGT 398
Db 61 AACCAAGCTTACCACTCCCTGTGTGTATTTGCTTCACTTCTCAATCCTTGAAGT 120
Qy 399 GGGCCGACATACCTACCTGCTCTCTACCTTTTGGCTTCCAGATGGCTTACCTTGGCT 458
Db 121 GGGCCGACATACCTACCTGCTCTCTACCTTTTGGCTTCCAGATGGCTTACCTTGGCT 180
Qy 459 GGATACATATACACGCGGCAAGTATGATATGATGAGCAATGCGACATTTGT 518
Db 181 GGATACATATACACGCGGCAAGTATGATATGATGAGCAATGCGACATTTGT 240
Qy 519 CTGACTTTGAAGCTGATTTGCTTGTGCTTGAATTTGACGAGGAGAAAGATCAGAT 578
Db 241 CTGACTTTGAAGCTGATTTGCTTGTGCTTGAATTTGACGAGGAGAAAGATCAGAT 300
Qy 579 TCCCTGTCTCTGAGCAACAAGATATGCGATACGTTGCTTCTTCCCTGCTGAAGT 638
Db 301 TCCCTGTCTCTGAGCAACAAGATATGCGATACGTTGCTTCTTCCCTGCTGAAGT 360
Qy 639 GGTGTTTCTCTCACTTCTATGAGGCTCTTGTGATGAGGCGCCAGTTCTCAATGATC 698
Db 361 GGTGTTTCTCTCACTTCTATGAGGCTCTTGTGATGAGGCGCCAGTTCTCAATGATC 420
Qy 699 TACATGAAGCTGTGAGGAGAGATGATGATGATGATGATGATGATGATGATGATG 758
Db 421 TACATGAAGCTGTGAGGAGAGATGATGATGATGATGATGATGATGATGATGATG 480
Qy 759 ATTCTGTCTCTCAAGGCTGTGAGTGTGAGGCTTTTCTTACCTAGTGGGCTACAC 818
Db 481 ATTCTGTCTCTCAAGGCTGTGAGTGTGAGGCTTTTCTTACCTAGTGGGCTACAC 540
Qy 819 ACCCCCAATCAGAGAGATATCTCTCACTGAGAGATATGATGATGATGATGATGAT 878
Db 541 ACCCCCAATCAGAGAGATATCTCTCACTGAGAGATATGATGATGATGATGATGAT 600
Qy 879 TTCCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 938

Db 601 TTCCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
Qy 938 TGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 998
Db 661 TGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
Qy 999 AAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 1058
Db 721 AAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 780
Qy 1059 CCCGCTTCTACTGAGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 1118
Db 781 CCCGCTTCTACTGAGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 840
Qy 1119 TACATCTTCAAGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 1178
Db 841 TACATCTTCAAGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 900
Qy 1179 CTATTCCTGAGGCTGTGAGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 1238
Db 901 CTATTCCTGAGGCTGTGAGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 960
Qy 1239 TTCCTCATTTGATTTGAGAAAGCAAGGCTGTGAGCAAGGCAAGGCAAGGCAAGG 1298
Db 961 TTCCTCATTTGATTTGAGAAAGCAAGGCTGTGAGCAAGGCAAGGCAAGGCAAGG 1020
Qy 1299 AGCAAGCTGCGCGCAATTAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCA 1358
Db 1021 AGCAAGCTGCGCGCAATTAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCA 1080
Qy 1359 CACTGAGCTTCTATGAGGCTGTGAGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 1418
Db 1081 CACTGAGCTTCTATGAGGCTGTGAGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 1140
Qy 1419 CTTAAGGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1478
Db 1141 CTTAAGGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200
Qy 1479 ATATTGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1538
Db 1201 ATATTGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1260
Qy 1539 TAAATCATTTCCCTGAGGCTGTGAGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 1598
Db 1261 TAAATCATTTCCCTGAGGCTGTGAGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 1320
Qy 1599 ACAGCACTCTTTGCGCGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 1658
Db 1321 ACAGCACTCTTTGCGCGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 1380
Qy 1659 TTCAAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 1718
Db 1381 TTCAAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 1440
Qy 1719 TTTGAGAAAGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 1778
Db 1441 TTTGAGAAAGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 1500
Qy 1779 ACCGCAACTCTTCTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGT 1838
Db 1501 ACCGCAACTCTTCTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGT 1560
Qy 1839 GGAATACAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 1898
Db 1561 GGAATACAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 1620
Qy 1899 CACCAAGTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCT 1958
Db 1621 CACCAAGTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCT 1680
Qy 1959 TTCCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2018

Db 1681 TCCCAAGTCTGGATTAACAGCGTGAACCCGCGCCCAAGGGAACACTT 1740
QY 2019 GTGGAGAGACAGAGGGGCTCACATCTCCCTCTGATTCCTCCATGCACTTCCCTTATC 2078
Db 1741 GTGGAGAGACAGAGGGGCTCACATCTCCCTCTGATTCCTCCATGCACTTCCCTTATC 1800
QY 2079 TCTCCCATCTAGCAGGAATCTATTGTGTGTTTCTTCTTCCGCC 2120
Db 1801 TCTCCCATCTAGCAGGAATCTATTGTGTGTTTCTTCTTCCGCC 1842

RESULT 3
US-09-954-531-1368
Sequence 1368, Application US/09954531
Patent No. US20020165180A1
GENERAL INFORMATION:
APPLICANT: Weaver, Zoe
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
TITLE OF INVENTION: Gene Sets
FILE REFERENCE: 689290-77
CURRENT APPLICATION NUMBER: US/09/954,531
CURRENT FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: US/60/233,133
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,009
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,034
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,509
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: US/60/234,567
PRIOR FILING DATE: 2000-09-22
NUMBER OF SEQ ID NOS: 1392
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1368
LENGTH: 1842
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-531-1368

Query Match 81.0%; Score 1842; DB 10; Length 1842;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1842; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 TACCTATCCACCTCTTCCATACCTTTACAGGCTCTCAATGGTTATTTAACTTTGA 338
Db 1 TACCTATCCACCTCTTCCATACCTTTACAGGCTCTCAATGGTTATTTAACTTTGA 60
QY 339 AACCAAGCTTACACCTCCGCTGTGATTTGCTTCAATCCATCCCTTCACTAATG 398
Db 61 AACCAAGCTTACACCTCCGCTGTGATTTGCTTCAATCCATCCCTTCACTAATG 120
QY 399 GGGCGACATCACTGCGCTCTCACTACCTTTTGTCCAGATGGCTACCTTCTGGCT 458
Db 121 GGGCGACATCACTGCGCTCTCACTACCTTTTGTCCAGATGGCTACCTTCTGGCT 180
QY 459 GGAATCTATTACACTGCGCAACCGGCACTAGATATCAAGTGAACATGCAATGTGTT 518
Db 181 GGAATCTATTACACTGCGCAACCGGCACTAGATATCAAGTGAACATGCAATGTGTT 240
QY 519 CTGACTTGAAGCTGATTTGTTGGCTGTGACTTGTGACGAGGAGGAAAGATCAGAT 578
Db 241 CTGACTTGAAGCTGATTTGTTGGCTGTGACTTGTGACGAGGAGGAAAGATCAGAT 300
QY 579 TCCCTGTCTCTGAGCAAGAAATATGCAATAGTGTGTTCCCTTCCCTGCTGGAAGTT 638
Db 301 TCCCTGTCTCTGAGCAAGAAATATGCAATAGTGTGTTCCCTTCCCTGCTGGAAGTT 360
QY 639 GCTGGTTTCTCTACTTCTATGGGCTTCTTGTGTAGGGCCCAAGTTCTCAATGAATAC 698
Db 361 GCTGGTTTCTCTACTTCTATGGGCTTCTTGTGTAGGGCCCAAGTTCTCAATGAATAC 420
QY 699 TACATGAAGCTGTGAGGAGAGCTGATTTGACATACAGGAAGATACCAACAGCATC 758

Db 421 TACATGAAGCTGTGAGGAGAGAGCTGATTTGACATACAGGAAGATACCAACACATC 480
QY 759 ATTCTGCTCTCAAGGCGCTGAGTCTGGGCTTTTCTTACATAGTGGCTTACACATGCTC 818
Db 481 ATTCTGCTCTCAAGGCGCTGAGTCTGGGCTTTTCTTACATAGTGGCTTACACATGCTC 540
QY 819 AGCCCCCATATCAGAGAAGATCTTCTCTACTGAAAGACTATAGAACCAACCCCTTCTG 878
Db 541 AGCCCCCATATCAGAGAAGATCTTCTCTACTGAAAGACTATAGAACCAACCCCTTCTG 600
QY 879 TTCGCTGATGATCATGTGATCTTGGGCAAGTTTGTCTGTATCAAAATATGTCACTGT 938
Db 601 TTCGCTGATGATCATGTGATCTTGGGCAAGTTTGTCTGTATCAAAATATGTCACTGT 660
QY 939 TGGCTGTCAAGAGAGATGATTTTGAACGGGCTGGGCTTCAATGCTTTGAAGAA 998
Db 661 TGGCTGTCAAGAGAGATGATTTTGAACGGGCTGGGCTTCAATGCTTTGAAGAA 720
QY 999 AAGGCAAGCAAGTGGGATGCTGTGCCAACATGAAGTGTGCTTTTGAACAAAC 1058
Db 721 AAGGCAAGCAAGTGGGATGCTGTGCCAACATGAAGTGTGCTTTTGAACAAAC 780
QY 1059 CCCCCTTCACTGGCACCAATGGCTCATTTCAACATCAACCAACGCTGGTGGCCGC 1118
Db 781 CCCCCTTCACTGGCACCAATGGCTCATTTCAACATCAACCAACGCTGGTGGCCGC 840
QY 1119 TACATTTCAACGACTCAAGTTCTTGGAAATTAAGAACTCTCTCAGGCTCTCTGTT 1178
Db 841 TACATTTCAACGACTCAAGTTCTTGGAAATTAAGAACTCTCTCAGGCTCTCTGTT 900
QY 1179 CTATTCCTGGCCCTCTGGACCGGCTGCACTAGATACCTGTCTTCCAGATGAA 1238
Db 901 CTATTCCTGGCCCTCTGGACCGGCTGCACTAGATACCTGTCTTCCAGATGAA 960
QY 1239 TTCCTATTTGTTATTTGGAAGACAGGCTGCAAGCTCATTTCAAGAGACCCCACTG 1298
Db 961 TTCCTATTTGTTATTTGGAAGACAGGCTGCAAGCTCATTTCAAGAGACCCCACTG 1020
QY 1299 AGCAAGCTGCGCGCAATTAAGTCTCTCCAGCCCTTCAATTTGGTGAACAGACATC 1358
Db 1021 AGCAAGCTGCGCGCAATTAAGTCTCTCCAGCCCTTCAATTTGGTGAACAGACATC 1080
QY 1359 CACTGCTCTTCAATGGTTACTCCAGATGCTCTTGTGCTTTCACGTGGGCAATATG 1418
Db 1081 CACTGCTCTTCAATGGTTACTCCAGATGCTCTTGTGCTTTCACGTGGGCAATATG 1140
QY 1419 CTTAAGTGTATTAATCCATCTATTCTTGGGCAATCTTCTCTGAGCTACTATTC 1478
Db 1141 CTTAAGTGTATTAATCCATCTATTCTTGGGCAATCTTCTCTGAGCTACTATTC 1200
QY 1479 ATATGCTTTATTTACAAAGCAATGTGCGCAAGGAAGAGAAGTTAAAGATGAA 1538
Db 1201 ATATGCTTTATTTACAAAGCAATGTGCGCAAGGAAGAGAAGTTAAAGATGAA 1260
QY 1539 TATATCAATTTCCCTGTGGGCTGTGGGCACTGTGTGCAAAATCAATGCTCCCTTTC 1598
Db 1261 TATATCAATTTCCCTGTGGGCTGTGGGCACTGTGTGCAAAATCAATGCTCCCTTTC 1320
QY 1599 ACAGCACTCTTTTGGCCCAAGAGAGAAATGAAAAGCAAGGAGGTGGAATGATGATC 1658
Db 1321 ACAGCACTCTTTTGGCCCAAGAGAGAAATGAAAAGCAAGGAGGTGGAATGATGATC 1380
QY 1659 TTCCAGCTGTGCTGTGCTGCCAGCAAGTCTTATTTGGGCAAGAGGGAATCTTTT 1718
Db 1381 TTCCAGCTGTGCTGTGCTGCCAGCAAGTCTTATTTGGGCAAGAGGGAATCTTTT 1440
QY 1719 TTTGAGAAAGGGCTTGTGCTTTGTACCGAGCTGGAATGCAAGTGGCGGGAATCTGAGCT 1778
Db 1441 TTTGAGAAAGGGCTTGTGCTTTGTACCGAGCTGGAATGCAAGTGGCGGGAATCTGAGCT 1500
QY 1779 ACCGCAACCTCAACTCTGTGGTTCAAGTATTTTCTGCTCAGCTCCCAAGTAGCTG 1838

```
Db 1501 ACCGCAACCTCCACCTCTGGGTTCAAGTATTTTCCTGCTCAGCTCCACCTCCAGTAGCTG 1560
QY 1839 GGAATACAGGACGAGCCACCATGCGCAGTAATTTTGTATTTTGTAGTAAAGGGATTT 1898
Db 1561 GGAATACAGGACGAGCCACCATGCGCAGTAATTTTGTATTTTGTAGTAAAGGGATTT 1620
QY 1899 CACCAAGTTGGCCAGGCTGGTCTCGAACTCTGACCGCAAGTATCCACCCGCTCCGCC 1958
Db 1621 CACCAAGTTGGCCAGGCTGGTCTCGAACTCTGACCGCAAGTATCCACCCGCTCCGCC 1680
QY 1959 TCCCAAGTCTGGGATTTACAGGCGTGAGCCACCGTCCCGGCCCAAGGGGAACTCTT 2018
Db 1681 TCCCAAGTCTGGGATTTACAGGCGTGAGCCACCGTCCCGGCCCAAGGGGAACTCTT 1740
QY 2019 GTGGAGGAGCAGAGGGGCTCATCTCCCTCTGATTTCCCATGACATTTGCTTATC 2078
Db 1741 GTGGAGGAGCAGAGGGGCTCATCTCTCCCTGATTTCCCATGACATTTGCTTATC 1800
QY 2079 TCTCCCATCTACGACGAGATCTATTTGTTTCTTCTGCC 2120
Db 1801 TCTCCCATCTACGACGAGATCTATTTGTTTCTTCTGCC 1842
```

RESULT 4

```
US-09-764-860-1135/c
; Sequence 1135, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1135
; LENGTH: 7461
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-860-1135
```

```
Query Match 31.1%; Score 708; DB 9; Length 7461;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1553 GGTGGCTGTGTGGGAGCTGTGTGAGAACTACTCTGTCCTTTTCAAGCACTCCTTTG 1612
Db 714 GGTGGCTGTGTGGGAGCTGTGTGAGAACTACTCTGTCCTTTTCAAGCACTCCTTTG 655
QY 1613 CCCGAGCAGAGAAATGAAAGCCAGGAGGTGAAAGTGCATGCTTCCAGCTGTGCT 1672
Db 654 CCCGAGCAGAGAAATGAAAGCCAGGAGGTGAAAGTGCATGCTTCCAGCTGTGCT 595
QY 1673 CTGCTGCCAGCAAGCTCTTCAATTTGGGGCCAAAGGGGAACTTTTGTGAGAAAGCGT 1732
Db 594 CTGCTGCCAGCAAGCTCTTCAATTTGGGGCCAAAGGGGAACTTTTGTGAGAAAGCGT 535
QY 1733 CTGCTTTGTACCCACGCTGGAATGCAATGCGGGATTCAGCTCACCGCAACTCCAC 1792
Db 534 CTGCTTTGTACCCACGCTGGAATGCAATGCGGGATTCAGCTCACCGCAACTCCAC 475
QY 1793 CTCCTGGGTTCAAGGATTTTCTGCTCAGCTCCCAAGTAGCTGGGAATTCAGGCAAG 1852
Db 474 CTCCTGGGTTCAAGGATTTTCTGCTCAGCTCCCAAGTAGCTGGGAATTCAGGCAAG 415
QY 1853 CCACCATGCCAGCTAATTTTGTATTTTGTAGTAAAGGGATTTCAACAGTTGGCCA 1912
Db 414 CCACCATGCCAGCTAATTTTGTATTTTGTAGTAAAGGGATTTCAACAGTTGGCCA 355
QY 1913 GGTGTGTCTGAACTCTCTGACCGCAAGTATCCACCGCTCCGCTCCCAAGTGTCTGG 1972
Db 354 GGTGTGTCTGAACTCTCTGACCGCAAGTATCCACCGCTCCGCTCCCAAGTGTCTGG 295
```

```
QY 1973 GATTACAGCGTGAGCCACCGTGGCCCGGCCCAAGGGGAAACTCTGTGGAGAGAGCAGA 2032
Db 294 GATTACAGCGTGAGCCACCGTGGCCCGGCCCAAGGGGAAACTCTGTGGAGAGAGCAGA 235
QY 2033 GGGGCTCAATCTCCCTCTGATTTCCCATGCACTTGTCTATCTCTCCCATCTAGC 2092
Db 234 GGGGCTCAATCTCCCTCTGATTTCCCATGCACTTGTCTATCTCTCCCATCTAGC 175
QY 2093 CAGGAATCTATTTGTTTCTTCTGCGCAATTTACTATGATTTGTATGTGCGCTACCA 2152
Db 174 CAGGAATCTATTTGTTTCTTCTGCGCAATTTACTATGATTTGTATGTGCGCTACCA 115
QY 2153 CCACCCCGCCCATGGGGGGGTGAGAGGGGGTGCAGAGGCCCTGGCTCCACTTTTCTA 2212
Db 114 CCACCCCGCCCATGGGGGGGTGAGAGGGGGTGCAGAGGCCCTGGCTCCACTTTTCTA 55
QY 2213 CTTGGAACGTGTATAGTAAATCACTTCTGTTTGTTCAGTTTTC 2260
Db 54 CTTGGAACGTGTATAGTAAATCACTTCTGTTTGTTCAGTTTTC 7
```

RESULT 5

```
US-09-764-904-91/c
; Sequence 91, Application US/09764904
; Patent No. US20020173454A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA122
; CURRENT APPLICATION NUMBER: US/09/764,904
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 91
; LENGTH: 7461
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-904-91
```

```
Query Match 31.1%; Score 708; DB 10; Length 7461;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1553 GGTGGCTGTGTGGGAGCTGTGTGAGAACTACTCTGTCCTTTTCAAGCACTCCTTTG 1612
Db 714 GGTGGCTGTGTGGGAGCTGTGTGAGAACTACTCTGTCCTTTTCAAGCACTCCTTTG 655
QY 1613 CCCGAGCAGAGAAATGAAAGCCAGGAGGTGAAAGTGCATGCTTCCAGCTGTGCT 1672
Db 654 CCCGAGCAGAGAAATGAAAGCCAGGAGGTGAAAGTGCATGCTTCCAGCTGTGCT 595
QY 1673 CTGCTGCCAGCAAGCTCTTCAATTTGGGGCCAAAGGGGAACTTTTGTGAGAAAGCGT 1732
Db 594 CTGCTGCCAGCAAGCTCTTCAATTTGGGGCCAAAGGGGAACTTTTGTGAGAAAGCGT 535
QY 1733 CTGCTTTGTACCCACGCTGGAATGCAATGCGGGATTCAGCTCACCGCAACTCCAC 1792
Db 534 CTGCTTTGTACCCACGCTGGAATGCAATGCGGGATTCAGCTCACCGCAACTCCAC 475
QY 1793 CTCCTGGGTTCAAGGATTTTCTGCTCAGCTCCCAAGTAGCTGGGAATTCAGGCAAG 1852
Db 474 CTCCTGGGTTCAAGGATTTTCTGCTCAGCTCCCAAGTAGCTGGGAATTCAGGCAAG 415
QY 1853 CCACCATGCCAGCTAATTTTGTATTTTGTAGTAAAGGGATTTCAACAGTTGGCCA 1912
Db 414 CCACCATGCCAGCTAATTTTGTATTTTGTAGTAAAGGGATTTCAACAGTTGGCCA 355
QY 1913 GGTGTGTCTGAACTCTCTGACCGCAAGTATCCACCGCTCCGCTCCCAAGTGTCTGG 1972
Db 354 GGTGTGTCTGAACTCTCTGACCGCAAGTATCCACCGCTCCGCTCCCAAGTGTCTGG 295
```

```
QY 1973 GATTACAGGCGTGAAGCCACCGTCCCGCCCAAGGGGAAACTCTTGTGGAGAGACAGA 2032
|||
Db 294 GATTACAGGCGTGAAGCCACCGTCCCGCCCAAGGGGAAACTCTTGTGGAGAGACAGA 235
QY 2033 GGGGCTCACATCTCCCTCTGATTTCCCATGACATTCCTTATCTCTCCCATCTAGC 2092
|||
Db 234 GGGGCTCACATCTCCCTCTGATTTCCCATGACATTCCTTATCTCTCCCATCTAGC 175
QY 2093 CAGGAATCTATTGTGTTTTCTTCTGCAATTACTATGATGTGATGTGCGCTACCA 2152
|||
Db 174 CAGGAATCTATTGTGTTTTCTTCTGCAATTACTATGATGTGATGTGCGCTACCA 115
QY 2153 CCACCCCCCATGGGGGGGTGAGAGGGGTGCAAGGCCCTGCTGCTCCACTTTTCTA 2212
|||
Db 114 CCACCCCCCATGGGGGGGTGAGAGGGGTGCAAGGCCCTGCTGCTCCACTTTTCTA 55
QY 2213 CCTTGAACCTGATATGATTAATAATCACTTCTGTTTGTTCAGTTTCTCA 2260
|||
Db 54 CCTTGAACCTGATATGATTAATAATCACTTCTGTTTGTTCAGTTTCTCA 7
```

```
RESULT 6
US-10-091-548-91/c
; Sequence 91, Application US/10091548
; Publication No. US20030049703A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA122C1
; CURRENT APPLICATION NUMBER: US/10/091,548
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ. ID NOS: 137
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ. ID NO 91
; LENGTH: 7461
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-091-548-91
```

```
Query Match 31.1%; Score 708; DB 14; Length 7461;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1553 GGTGGCTGTGGCGGAGCTGTGCGAAGAACTACTCTCTTTCACAGACTCTTTG 1612
|||
Db 714 GGTGGCTGTGGCGGAGCTGTGCGAAGAACTACTCTCTTTCACAGACTCTTTG 655
QY 1613 CCCGAGCAGAGAAATGAAAAGCCAGGAGGTGGAAGATCGATCTTCCAGCTGTGCT 1672
|||
Db 654 CCCGAGCAGAGAAATGAAAAGCCAGGAGGTGGAAGATCGATCTTCCAGCTGTGCT 595
QY 1673 CTGTGTCAGCAAGTCTTCAATTTGGGCCAAGAGGGAACCTTTTTTGGAGAAGCGT 1732
|||
Db 594 CTGTGTCAGCAAGTCTTCAATTTGGGCCAAGAGGGAACCTTTTTTGGAGAAGCGT 535
QY 1733 CTGTGTCAGCAAGTCTTCAATTTGGGCCAAGAGGGAACCTTTTTTGGAGAAGCGT 1792
|||
Db 534 CTGTGTCAGCAAGTCTTCAATTTGGGCCAAGAGGGAACCTTTTTTGGAGAAGCGT 475
QY 1793 CTCTGGGTTCAAGTATTTTCCGCTCAGCCTCCCAAGTATGGAATATACAGGACG 1852
|||
Db 474 CTCTGGGTTCAAGTATTTTCCGCTCAGCCTCCCAAGTATGGAATATACAGGACG 415
QY 1853 CCACCATGCCAGCAATTTTGTATTTTCAAGTAAACGGGATTTTCAACAGTTGGCA 1912
|||
Db 414 CCACCATGCCAGCAATTTTGTATTTTCAAGTAAACGGGATTTTCAACAGTTGGCA 355
QY 1913 GGTGTGTCTGAATCTCTGACCGCAAGTATCCACCGGCTCTCCGCTCCCAAGTGTGG 1972
|||
Db 354 GGTGTGTCTGAATCTCTGACCGCAAGTATCCACCGGCTCTCCGCTCCCAAGTGTGG 295
QY 1973 GATTACAGGCGTGAAGCCACCGTCCCGCCCAAGGGGAAACTCTTGTGGAGAGACAGA 2032
```

```
Db 294 GATTACAGGCGTGAAGCCACCGTCCCGCCCAAGGGGAAACTCTTGTGGAGAGACAGA 235
|||
QY 2033 GGGGCTCACATCTCCCTCTGATTTCCCATGACATTCCTTATCTCTCCCATCTAGC 2092
|||
Db 234 GGGGCTCACATCTCCCTCTGATTTCCCATGACATTCCTTATCTCTCCCATCTAGC 175
QY 2093 CAGGAATCTATTGTGTTTTCTTCTGCAATTACTATGATGTGATGTGCGCTACCA 2152
|||
Db 174 CAGGAATCTATTGTGTTTTCTTCTGCAATTACTATGATGTGATGTGCGCTACCA 115
QY 2153 CCACCCCCCATGGGGGGGTGAGAGGGGTGCAAGGCCCTGCTGCTCCACTTTTCTA 2212
|||
Db 114 CCACCCCCCATGGGGGGGTGAGAGGGGTGCAAGGCCCTGCTGCTCCACTTTTCTA 55
QY 2213 CCTTGAACCTGATATGATTAATAATCACTTCTGTTTGTTCAGTTTCTCA 2260
|||
Db 54 CCTTGAACCTGATATGATTAATAATCACTTCTGTTTGTTCAGTTTCTCA 7
```

```
RESULT 7
US-10-074-095-1135/c
; Sequence 1135, Application US/10074095
; Publication No. US20030077704A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008C1
; CURRENT APPLICATION NUMBER: US/10/074,095
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 09/764,860
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/225,757
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/226,868
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/216,647
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,267
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/216,880
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/225,270
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/251,869
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/235,834
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/234,274
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/234,223
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 60/228,924
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/224,518
```

```

PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/236,369
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/224,519
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/220,964
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/241,809
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/249,299
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/236,327
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/241,785
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/244,617
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: 60/225,268
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/236,368
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/251,856
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/251,868
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/229,344
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/234,997
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: 60/229,343
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,345
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,287
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,513
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/231,413
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/229,509
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/236,367
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/237,039
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,038
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/236,370
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/236,802
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,037
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,040
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/240,960
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/239,935
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/239,937
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/241,787
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,474
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/246,532
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/249,216
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,210
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/226,681
PRIOR FILING DATE: 2000-08-22

```

```

PRIOR APPLICATION NUMBER: 60/225,759
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/225,213
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/227,182
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,214
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/235,836
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: 60/230,438
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/215,135
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: 60/225,266
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/249,218
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,208
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,213
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,212
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,207
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,245
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,244
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,217
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,211
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,215
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,264
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,214
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,297
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/232,400
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/231,242
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,081
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,080
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/231,414
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/231,244
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/233,064
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/233,063
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,397
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,399
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,401
PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/241,808
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,826
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,786
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/241,221
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,475

```



```
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match      31.1%; Score 708; DB 14; Length 7461;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1553 GGTGGCCCTGTGGCGGAGTGTGAGAAACTACTGCTCTCTCTTTTCACAGCACTCTCTTTG 1612
DB 714 GGTGGCCCTGTGGCGGAGTGTGAGAAACTACTGCTCTCTCTTTTCACAGCACTCTCTTTG 655
QY 1613 CCCGAGGAGAGAAATGGAAGAGCGAGGAGTGAAGATGCTTCCAGCTGAGCGCT 1672
DB 654 CCCGAGGAGAGAAATGGAAGAGCGAGGAGTGAAGATGCTTCCAGCTGAGCGCT 595
QY 1673 CTGCTGCGACGCAAGTCTTCAATTTGGGGCAAGGGGAACTTTTGGAGAAGCGCT 1732
DB 594 CTGCTGCGACGCAAGTCTTCAATTTGGGGCAAGGGGAACTTTTGGAGAAGCGCT 535
QY 1733 CTGCTGCTGTACCCAGCTGGAATGAGTGGGGATCTCAAGCTCAGCGCAACTCTTCAC 1792
DB 534 CTGCTGCTGTACCCAGCTGGAATGAGTGGGGATCTCAAGCTCAGCGCAACTCTTCAC 475
QY 1793 CTGCTGGGTCAAGTATTTCTGCTCTCAGCCCTCCCAAGTACTGAGGAAATAGAGCAGC 1852
DB 474 CTGCTGGGTCAAGTATTTCTGCTCTCAGCCCTCCCAAGTACTGAGGAAATAGAGCAGC 415
QY 1853 CCACACATGCCACGCTAATTTTGTATTTTCAAGTAAAGCGGATTTTCAACACGTTGGCCA 1912
DB 414 CCACACATGCCACGCTAATTTTGTATTTTCAAGTAAAGCGGATTTTCAACACGTTGGCCA 355
QY 1913 GGTGCTGTCTGAATCTCTGACCCGCAAGTATCCACCCGCTCTCGCTCCCAAGTGTGG 1972
DB 354 GGTGCTGTCTGAATCTCTGACCCGCAAGTATCCACCCGCTCTCGCTCCCAAGTGTGG 295
QY 1973 GATTACAGGGGTGAGCAGCGTGCCTGGCCCAAGGGGAAACTTGTGGAGAGAGAGA 2032
DB 294 GATTACAGGGGTGAGCAGCGTGCCTGGCCCAAGGGGAAACTTGTGTGGAGAGAGAGA 235
QY 2033 GGGGCTCACATCTCCCTCTGATTTCCGCCATGACATGCTTATCTCTCCCATCTAGC 2092
DB 234 GGGGCTCACATCTCCCTCTGATTTCCGCCATGACATGCTTATCTCTCCCATCTAGC 175
QY 2093 CAGGATCTAATGTGTTTTCTTCTGCAATTTAATGATGATGTAATGTCGCGCTACCA 2152
DB 174 CAGGATCTAATGTGTTTTCTTCTGCAATTTAATGATGATGTAATGTCGCGCTACCA 115
QY 2153 CCACCCCCCGCATGGGGGGGAGAGAGGGGTGCAAGGCCCTGCTGCTCAGCTTTTCTA 2212
DB 114 CCACCCCCCGCATGGGGGGGAGAGAGGGGTGCAAGGCCCTGCTGCTCAGCTTTTCTA 55
QY 2213 CCTTGAAGCTGATTAATAAATCACTTCTGTTTGTTCAGTTTCTA 2260
DB 54 CCTTGAAGCTGATTAATAAATCACTTCTGTTTGTTCAGTTTCTA 7

RESULT 8
US-09-908-711-158/c
; Sequence 158, Application US/09908711
; Patent No. US20020045230A1
; GENERAL INFORMATION:
; APPLICANT: Roese et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA128
; CURRENT APPLICATION NUMBER: US/09/908,711
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US01/01360
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,867
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01344
; PRIOR FILING DATE: 2001-01-17
```

```
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01345
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,888
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01329
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,905
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01354
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,891
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01339
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,869
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01340
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,874
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01334
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,898
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01320
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,853
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01349
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,902
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01239
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,870
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01348
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,882
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01347
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,896
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01307
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,864
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01341
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,856
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01336
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/764,868
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: US01/01312
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatencIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 32174
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```


NAME/KEY: SITE	LOCATION: (29356)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29357)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29358)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29359)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29360)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29361)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29362)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29363)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29364)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29365)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29366)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29367)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29368)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29369)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29370)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29371)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29372)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29373)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29374)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29375)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29376)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29377)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29378)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29379)	OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE	LOCATION: (29380)	OTHER INFORMATION: n equals a,t,g, or c

```

LOCATION: (29380)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29381)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29382)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29383)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29384)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29385)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29386)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29387)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29388)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29389)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29390)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29391)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29392)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29393)

```

Query Match	31.1%;	Score 708;	DB 9;	length 32174;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 708;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1553	GGTGGCGCTGGCGGGATCTGGAGCAGAAACTACTGCTCTCCCTTTTCAACAGACTCCTTGG	1612
Db	6043	GGTGGCGCTGGCGGGATCTGGAGCAGAAACTACTGCTCTCCCTTTTCAACAGACTCCTTGG	5994
QY	1613	CCCCAGAGCAGAGAAATGAAAAGCCAGGAGAGTGGAAATGCAATGCTTCCAGCTGTGCTCT	1672
Db	5983	CCCCAGAGCAGAGAAATGAAAAGCCAGGAGAGTGGAAATGCAATGCTTCCAGCTGTGCTCT	5924
QY	1673	CTGCTGCCAGCMAAGTCTTCAATTTGGGGCCAAAGGGGAAACTTTTTTTGGAGAGGCGT	1732
Db	5923	CTGCTGCCAGCMAAGTCTTCAATTTGGGGCCAAAGGGGAAACTTTTTTTGGAGAGGCGT	5864
QY	1733	CTTGGCTTTGTACCCACGCTGGAAATGCAAGTGGGGGAACTCAGCTCACCCCAACCTTCCAC	1792
Db	5863	CTTGGCTTTGTACCCACGCTGGAAATGCAAGTGGGGGAACTCAGCTCACCCCAACCTTCCAC	5804
QY	1793	CTCTCTGGGTTCAAGTATTTTTCCTGCTCAGCCTCCCAAGTAGCTGGGAAATACAGGACAG	1852
Db	5803	CTCTCTGGGTTCAAGTATTTTTCCTGCTCAGCCTCCCAAGTAGCTGGGAAATACAGGACAG	5744
QY	1853	CCACCATGCCAGCTAATTTTGTATTTTCAGTAGAAACGGGATTTTCAACAAGTGTGGCCA	1912
Db	5743	CCACCATGCCAGCTAATTTTGTATTTTCAGTAGAAACGGGATTTTCAACAAGTGTGGCCA	5684
QY	1913	GGGCTGGCTGAAGTCCCTGACCCGCAAGTAGATCCACCCGCTCCGCTCCCAAAAGTGTGG	1972
Db	5683	GGGCTGGCTGAAGTCCCTGACCCGCAAGTAGATCCACCCGCTCCGCTCCCAAAAGTGTGG	5624


```
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29392)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29393)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29394)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29395)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29396)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29397)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29398)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29399)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29400)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29401)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29402)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29403)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29404)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29405)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29406)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29407)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29408)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29409)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29410)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29411)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29412)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29413)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29414)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29415)
OTHER INFORMATION: n equals a,t,g, or c
```

```
NAME/KEY: SITE
LOCATION: (29416)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE

Query Match      31.1%; Score 708; DB 9; Length 32174;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1553 GGTGGCTGTGCGGAGCTGTGACAGAACTACTGCTCTCCCTTTTCACAGCACTCTTTG 1612
DB GGTGGCTGTGCGGAGCTGTGACAGAACTACTGCTCTCCCTTTTCACAGCACTCTTTG 5984

1613 CCCGACAGAGATGAGAAAGCCAGGAGGTGAGATGATGATCTTCAGCTGAGCT 1672
DB CCCGACAGAGATGAGAAAGCCAGGAGGTGAGATGATGATCTTCAGCTGAGCT 5924

1673 CTGCTGCGCAGGCAAGTCTTCAATTTGGGGCCAAAGGGAACTTTTGTGAGAGCGT 1732
DB CTGCTGCGCAGGCAAGTCTTCAATTTGGGGCCAAAGGGAACTTTTGTGAGAGCGT 5864

1733 CTGCTTTGTCAACCGCTGGAATGCAATGCGGGATCTCACTCAACGCACTCCAC 1792
DB CTGCTTTGTCAACCGCTGGAATGCAATGCGGGATCTCACTCAACGCACTCCAC 5804

1793 CTCCGGGGTCAAGATTTTCTGCTGAGCTCCAGCTCCCAAGTAGCTGGAAATCAGGACAG 1852
DB CTCCGGGGTCAAGATTTTCTGCTGAGCTCCAGCTCCCAAGTAGCTGGAAATCAGGACAG 5744

1853 CCACCATGCCAGCTAATTTTGTATTTTCAATAGTAAGCGGATTTCAACCACTGGCCA 1912
DB CCACCATGCCAGCTAATTTTGTATTTTCAATAGTAAGCGGATTTCAACCACTGGCCA 5684

1913 GGCTGTCTCGAATCTCTGACCGCAAGTATCAACCGCTCCGCTCCCAAGTCTGG 1972
DB GGCTGTCTCGAATCTCTGACCGCAAGTATCAACCGCTCCGCTCCCAAGTCTGG 5624

1973 GATTACAGGCGTGACGACCGCGGCGCCCAAGGGGAACTCTTGTGGAGAGACAGA 2032
DB GATTACAGGCGTGACGACCGCGGCGCCCAAGGGGAACTCTTGTGGAGAGACAGA 5564

2033 GGGGCTCAGATCTCCCTCTGATTCGCCCATGACATTCCTATCTCTCCCATCTAC 2092
DB GGGGCTCAGATCTCCCTCTGATTCGCCCATGACATTCCTATCTCTCCCATCTAC 5504

2093 CAGGAATCTATTTGTTTTTTCTTCTGCCAATTACTATGATTTGTATGTCGCTACCA 2152
DB CAGGAATCTATTTGTTTTTTCTTCTGCCAATTACTATGATTTGTATGTCGCTACCA 5444

2153 CCACCCCGCCCATGCGGGGGGTGAGAGGGGGTGCAGAGCCGCTGCTCCATTCTTTCTA 2212
DB CCACCCCGCCCATGCGGGGGGTGAGAGGGGGTGCAGAGCCGCTGCTCCATTCTTTCTA 5384

2213 CCTTGAATCTATTTAGATAAATCATCTTCTTTTGTTCAGTTTCTA 2260
DB CCTTGAATCTATTTAGATAAATCATCTTCTTTTGTTCAGTTTCTA 5336

RESULT 10
US-09-764-877-2645/c
Sequence 2645, Application US/09764877
Patent No. US20020147140A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC005
CURRENT APPLICATION NUMBER: US/09/764,877
PRIORITY FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2645
LENGTH: 32174
```

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

1	LOCATION: (29379)	n	equals	a,t,g,	or c
2	OTHER INFORMATION:				
3	NAME/KEY: SITE				
4	LOCATION: (29380)	n	equals	a,t,g,	or c
5	OTHER INFORMATION:				
6	NAME/KEY: SITE				
7	LOCATION: (29381)	n	equals	a,t,g,	or c
8	OTHER INFORMATION:				
9	NAME/KEY: SITE				
10	LOCATION: (29382)	n	equals	a,t,g,	or c
11	OTHER INFORMATION:				
12	NAME/KEY: SITE				
13	LOCATION: (29383)	n	equals	a,t,g,	or c
14	OTHER INFORMATION:				
15	NAME/KEY: SITE				
16	LOCATION: (29384)	n	equals	a,t,g,	or c
17	OTHER INFORMATION:				
18	NAME/KEY: SITE				
19	LOCATION: (29385)	n	equals	a,t,g,	or c
20	OTHER INFORMATION:				
21	NAME/KEY: SITE				
22	LOCATION: (29386)	n	equals	a,t,g,	or c
23	OTHER INFORMATION:				
24	NAME/KEY: SITE				
25	LOCATION: (29387)	n	equals	a,t,g,	or c
26	OTHER INFORMATION:				
27	NAME/KEY: SITE				
28	LOCATION: (29388)	n	equals	a,t,g,	or c
29	OTHER INFORMATION:				
30	NAME/KEY: SITE				
31	LOCATION: (29389)	n	equals	a,t,g,	or c
32	OTHER INFORMATION:				
33	NAME/KEY: SITE				
34	LOCATION: (29390)	n	equals	a,t,g,	or c
35	OTHER INFORMATION:				
36	NAME/KEY: SITE				
37	LOCATION: (29391)	n	equals	a,t,g,	or c
38	OTHER INFORMATION:				
39	NAME/KEY: SITE				
40	LOCATION: (29392)	n	equals	a,t,g,	or c
41	OTHER INFORMATION:				
42	NAME/KEY: SITE				
43	LOCATION: (29393)	n	equals	a,t,g,	or c
44	OTHER INFORMATION:				
45	NAME/KEY: SITE				
46	LOCATION: (29394)	n	equals	a,t,g,	or c
47	OTHER INFORMATION:				
48	NAME/KEY: SITE				
49	LOCATION: (29395)	n	equals	a,t,g,	or c
50	OTHER INFORMATION:				
51	NAME/KEY: SITE				
52	LOCATION: (29396)	n	equals	a,t,g,	or c
53	OTHER INFORMATION:				
54	NAME/KEY: SITE				
55	LOCATION: (29397)	n	equals	a,t,g,	or c
56	OTHER INFORMATION:				
57	NAME/KEY: SITE				
58	LOCATION: (29398)	n	equals	a,t,g,	or c
59	OTHER INFORMATION:				
60	NAME/KEY: SITE				
61	LOCATION: (29399)	n	equals	a,t,g,	or c
62	OTHER INFORMATION:				
63	NAME/KEY: SITE				
64	LOCATION: (29400)	n	equals	a,t,g,	or c
65	OTHER INFORMATION:				
66	NAME/KEY: SITE				
67	LOCATION: (29401)	n	equals	a,t,g,	or c
68	OTHER INFORMATION:				
69	NAME/KEY: SITE				
70	LOCATION: (29402)	n	equals	a,t,g,	or c
71	OTHER INFORMATION:				
72	NAME/KEY: SITE				
73	LOCATION: (29403)	n	equals	a,t,g,	or c


```

; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29416)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE

```

Query Match	31.1%	Score 708;	DB 10;	Length 32174;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 708; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

OY	1553	GATGACCTGTGGGGACCTGGTGCAGAAATCTACGCTCCCTTTTCACAGCACTCTTTG	161
Db	6043	GGTGGCCTGTGGGGACCTGGTGCAGAAATCTACGCTCCCTTTTCACAGCACTCTTTG	5984
OY	1613	CCCCAGACAGAAATGAAAAAGCCAGGAGGTGAGATGCATGCTTCCACTGTGCT	1677
Db	5983	CCCCAGACAGAGAAATGAAAAAGCCAGGAGGTGAGATGCATGCTTCCACTGTGCT	5922
OY	1673	CTGCTGGCAGCCCAATCTTCAATTTGGGGCCAAAGGGAAACTTTTTTGGAAAGGCT	1732
Db	5923	CTGCTGGCAGCCCAATCTTCAATTTGGGGCCAAAGGGAAACTTTTTTGGAAAGGCT	5862
OY	1733	CTTGCTTTGTCAACCCACGCTGGAATGCAGTGGCGGGATCTCAGCTCACCGCAACCTCCAC	1792
Db	5863	CTTGCTTTGTCAACCCACGCTGGAATGCAGTGGCGGGATCTCAGCTCACCGCAACCTCCAC	5802
OY	1793	CTCCCTGGGTCAAGTATTTTCTGCTCAGCCTCCCAAGTACGTGGGAATACAGCACG	1852
Db	5803	CTCCCTGGGTCAAGTATTTTCTGCTCAGCCTCCCAAGTACGTGGGAATACAGCACG	5744
OY	1853	CCACCATCCCGAGCTAAATTTTGTATTTTCATGAGAAAGGGAATTCACAGTGGGCA	1912
Db	5743	CCACCATCCCGAGCTAAATTTTGTATTTTCATGAGAAAGGGAATTCACAGTGGGCA	5682
OY	1913	GGCTGGTCTCGAATCTCTGACCGCAAGTATCCACCCGCTTCGCGCTCCAAAGTGTGG	1972
Db	5683	GGCTGGTCTCGAATCTCTGACCGCAAGTATCCACCCGCTTCGCGCTCCAAAGTGTGG	5622
OY	1973	GATTACAGGGCGTAGACCAACGAGCCCGGCCAAAGGGGAAACTCTGTGGGAGGAGCA	2032
Db	5623	GATTACAGGGCGTAGACCAACGAGCCCGGCCAAAGGGGAAACTCTGTGGGAGGAGCA	5562
OY	2033	GGGGCTCACTCTCCCTCTGATTCGCCCATGCACATTTGCCCTTATCTCTCCCATCTAGC	2092
Db	5563	GGGGCTCACTCTCCCTCTGATTCGCCCATGCACATTTGCCCTTATCTCTCCCATCTAGC	5502
OY	2093	CAGGAATCTATTTGGTTTTTCTTCTGACCAATTTACTATGATGTGTGATGTGCGGTACCA	2152
Db	5503	CAGGAATCTATTTGGTTTTTCTTCTGACCAATTTACTATGATGTGTGATGTGCGGTACCA	5442
OY	2153	CCACCCCCCCATGGGGGGGTGGAAGAGGGTGCAGAGCCCTGACCTGCATCTTTTCTA	2212
Db	5443	CCACCCCCCCATGGGGGGGTGGAAGAGGGTGCAGAGCCCTGACCTGCATCTTTTCTA	5382
OY	2213	CCTTGGACTGTATTAAGTAAAAATCACTTTCTGTGTTCACGTTTTTCA 2260	
Db	5383	CCTTGGAACTGTATTAAGTAAAAATCACTTGTGTTCACGTTTTTCA 5336	

```

RESULT 12
US-09-764-904/90/C
; Sequence 90, Application US/09764904
; Patent No. US20020173454A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA122
; CURRENT APPLICATION NUMBER: US/09/764, 904
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 90

```

1	LENGTH: 32174	
2	TYPE: DNA	
3	ORGANISM: Homo sapiens	
4	FEATURE:	
5	NAME/KEY: SITE	
6	LOCATION: (29356)	n equals a,t,g, or c
7	OTHER INFORMATION:	
8	NAME/KEY: SITE	
9	LOCATION: (29357)	n equals a,t,g, or c
10	OTHER INFORMATION:	
11	NAME/KEY: SITE	
12	LOCATION: (29358)	n equals a,t,g, or c
13	OTHER INFORMATION:	
14	NAME/KEY: SITE	
15	LOCATION: (29359)	n equals a,t,g, or c
16	OTHER INFORMATION:	
17	NAME/KEY: SITE	
18	LOCATION: (29360)	n equals a,t,g, or c
19	OTHER INFORMATION:	
20	NAME/KEY: SITE	
21	LOCATION: (29361)	n equals a,t,g, or c
22	OTHER INFORMATION:	
23	NAME/KEY: SITE	
24	LOCATION: (29362)	n equals a,t,g, or c
25	OTHER INFORMATION:	
26	NAME/KEY: SITE	
27	LOCATION: (29363)	n equals a,t,g, or c
28	OTHER INFORMATION:	
29	NAME/KEY: SITE	
30	LOCATION: (29364)	n equals a,t,g, or c
31	OTHER INFORMATION:	
32	NAME/KEY: SITE	
33	LOCATION: (29365)	n equals a,t,g, or c
34	OTHER INFORMATION:	
35	NAME/KEY: SITE	
36	LOCATION: (29366)	n equals a,t,g, or c
37	OTHER INFORMATION:	
38	NAME/KEY: SITE	
39	LOCATION: (29367)	n equals a,t,g, or c
40	OTHER INFORMATION:	
41	NAME/KEY: SITE	
42	LOCATION: (29368)	n equals a,t,g, or c
43	OTHER INFORMATION:	
44	NAME/KEY: SITE	
45	LOCATION: (29369)	n equals a,t,g, or c
46	OTHER INFORMATION:	
47	NAME/KEY: SITE	
48	LOCATION: (29370)	n equals a,t,g, or c
49	OTHER INFORMATION:	
50	NAME/KEY: SITE	
51	LOCATION: (29371)	n equals a,t,g, or c
52	OTHER INFORMATION:	
53	NAME/KEY: SITE	
54	LOCATION: (29372)	n equals a,t,g, or c
55	OTHER INFORMATION:	
56	NAME/KEY: SITE	
57	LOCATION: (29373)	n equals a,t,g, or c
58	OTHER INFORMATION:	
59	NAME/KEY: SITE	
60	LOCATION: (29374)	n equals a,t,g, or c
61	OTHER INFORMATION:	
62	NAME/KEY: SITE	
63	LOCATION: (29375)	n equals a,t,g, or c
64	OTHER INFORMATION:	
65	NAME/KEY: SITE	
66	LOCATION: (29376)	n equals a,t,g, or c
67	OTHER INFORMATION:	
68	NAME/KEY: SITE	
69	LOCATION: (29377)	n equals a,t,g, or c
70	OTHER INFORMATION:	
71	NAME/KEY: SITE	
72	LOCATION: (29378)	n equals a,t,g, or c
73	OTHER INFORMATION:	

```

NAME/KEY: SITE
LOCATION: (29379)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29380)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29381)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29382)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29383)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29384)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29385)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29386)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29387)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29388)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29389)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29390)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29391)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29392)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29393)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29394)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29395)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29396)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29397)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29398)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29400)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29401)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29402)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE

```

```

LOCATION: (29403)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29404)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29405)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29406)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29407)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29408)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29409)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29410)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29411)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29412)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29413)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29414)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29415)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29416)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE

```

Query Match 31.1%; Score 708; DB 10; Length 32174;
 Best Local Similarity 100.0%; Fred. No. 0;
 Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1553 GGTGGCTGTGCGGAGCTGTGAGAACTACTGTCCTTTTCACAGCACTCCCTTG 1612
DB 6043 GGTGGCTGTGCGGAGCTGTGAGAACTACTGTCCTTTTCACAGCACTCCCTTG 5984
QY 1613 CCCGAGCAGAGATGAGAAAGCCAGGAGGTGAGATGATGATGCTTCCAGCTGTGCT 1672
DB 5983 CCCGAGCAGAGAAAGGAAAGCCAGGAGGTGAGATGATGATGCTTCCAGCTGTGCT 5924
QY 1673 CTGCTGCCAGCCAGATCTTATTTGGGCGCAAGGGGAACTTTTGGAGAAAGGCT 1732
DB 5923 CTGCTGCCAGCCAGATCTTATTTGGGCGCAAGGGGAACTTTTGGAGAAAGGCT 5864
QY 1733 CTGCTTTGTCAACCCAGCTGATGATGAGGGGAGCTTCAGCTCAACGCACTTCAC 1792
DB 5863 CTGCTTTGTCAACCCAGCTGATGATGAGGGGAGCTTCAGCTCAACGCACTTCAC 5804
QY 1793 CTCTGGGTTCAAGATATTTCTGCTCAGCTCCCAAGTACCTGGAGATACAGGCACG 1852
DB 5803 CTCTGGGTTCAAGATATTTCTGCTCAGCTCCCAAGTACCTGGAGATACAGGCACG 5744
QY 1853 CCACCATGCCAGCTAATTTTGTATTTCAAGTAAAGGGATTTTACCAAGTTGGCCA 1912
DB 5743 CCACCATGCCAGCTAATTTTGTATTTCAAGTAAAGGGATTTTACCAAGTTGGCCA 5684
QY 1913 GGCTGCTCGAATCTCTGACGCAAGTATCAACCGGCTCGCTCCCAAGTCTGG 1972
DB 5684 GGCTGCTCGAATCTCTGACGCAAGTATCAACCGGCTCGCTCCCAAGTCTGG 1912

```


Db 5683 GGGTGTCTCGAACTCTCTGACCCGCAAGTATCCACCCGCTCCGCTCCCAAGTGTGG 5624
 Qy 1973 GATTACAGGGGTGAGGACACCGTGGCCGCAAGGGGAACTCTGTGGAGAGACCA 2032
 Db 5623 GATTACAGGGGTGAGGACACCGTGGCCGCAAGGGGAACTCTGTGGAGAGACCA 5564
 Qy 2033 GGGGCTCACATCTCCCTCTGATTCCTCCCATGACATTCCTTATCTCCCATCTAGC 2092
 Db 5563 GGGGCTCACATCTCCCTCTGATTCCTCCCATGACATTCCTTATCTCCCATCTAGC 5504
 Qy 2093 CAGGAATCTATGTGTGTCTCTCTGCAATTTACTATGATGTGTATGTCGCTACCA 2152
 Db 5503 CAGGAATCTATGTGTGTCTCTCTGCAATTTACTATGATGTGTATGTCGCTACCA 5444
 Qy 2153 CCACCCCTCCCATGAGGGGGTGGAGAGGGGTGCAAGCCCTGCTGCTCCACTTTTCA 2212
 Db 5443 CCACCCCTCCCATGAGGGGGTGGAGAGGGGTGCAAGCCCTGCTGCTCCACTTTTCA 5384
 Qy 2213 CCTTGGAAGTGTATGATTAATAATCACTTCTGTTGTGAGTTTTC 2260
 Db 5383 CCTTGGAAGTGTATGATTAATAATCACTTCTGTTGTGAGTTTTC 5336

RESULT 13
 US-09-764-891-6480/c
 ; Sequence 6480, Application US/09764891
 ; Publication No. US20030077808A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006
 ; CURRENT APPLICATION NUMBER: US/09/764,891
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 10231
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 6480
 ; LENGTH: 32174
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (29356)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29357)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29358)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29359)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29360)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29361)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29362)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29363)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29364)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29365)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29366)

OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29367)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29368)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29369)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29370)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29371)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29372)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29373)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29374)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29375)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29376)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29377)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29378)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29379)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29380)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29381)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29382)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29383)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29384)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29385)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29386)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29387)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29388)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29389)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (29390)
 ; OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE
LOCATION: (29391) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29392) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29393) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29394) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29395) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29396) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29397) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29398) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29399) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29400) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29401) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29402) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29403) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29404) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29405) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29406) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29407) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29408) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29409) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29410) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29411) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29412) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29413) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29414) n equals a,t,g, or c
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE

LOCATION: (29415)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (29416)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE

Query Match 31.1%; Score 708; DB 11; Length 32174;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1553 GGTGGCTGTGGGGGAGCTGTGAGAACTACCTGCTCCCTTTTACAGCACTCTTTG 1612
6043 GGTGGCTGTGGGGGAGCTGTGAGAACTACCTGCTCCCTTTTACAGCACTCTTTG 5984
1613 CCCGAGCAGAGAAATGAGAAAGCCAGGAGGTGAGAAATGATGCTTCCAGCTGTGCT 1672
5983 CCCGAGCAGAGAAATGAGAAAGCCAGGAGGTGAGAAATGATGCTTCCAGCTGTGCT 5924
1673 CTGCTGCGCAGCAAGCTTCACTTTGGGGCCAAAGGGGAACTTTTGGAGAGGCGT 1732
5923 CTGCTGCGCAGCAAGCTTCACTTTGGGGCCAAAGGGGAACTTTTGGAGAGGCGT 5864
1733 CTGCTGCGCAGCAAGCTTCACTTTGGGGCCAAAGGGGAACTTTTGGAGAGGCGT 1792
5863 CTGCTGCGCAGCAAGCTTCACTTTGGGGCCAAAGGGGAACTTTTGGAGAGGCGT 5804
1793 CTGCTGCGCAGCAAGCTTCACTTTGGGGCCAAAGGGGAACTTTTGGAGAGGCGT 1852
5803 CTGCTGCGCAGCAAGCTTCACTTTGGGGCCAAAGGGGAACTTTTGGAGAGGCGT 5744
1853 CCACATGCGCCAGCTAATTTTGTATTTTCAATAGAAAGGAACTTTTACAGCTGGGCA 1912
5743 CCACATGCGCCAGCTAATTTTGTATTTTCAATAGAAAGGAACTTTTACAGCTGGGCA 5684
1913 GGTGGCTGTGGGGGAGCTGTGAGAACTACCTGCTCCCTTTTACAGCACTCTTTG 1972
5683 GGTGGCTGTGGGGGAGCTGTGAGAACTACCTGCTCCCTTTTACAGCACTCTTTG 5624
1973 GATTACAGCGGTGAGCCAGCTGCGCGGCCCAAGGGGAACTTTTGGAGAGGCGT 2032
5623 GATTACAGCGGTGAGCCAGCTGCGCGGCCCAAGGGGAACTTTTGGAGAGGCGT 5564
2033 GGGGCTCACATCTCCCTGATTCCTCCCATGCACTGCTTATCTCCCATCTAGC 2092
5563 GGGGCTCACATCTCCCTGATTCCTCCCATGCACTGCTTATCTCCCATCTAGC 5504
2093 CAGGATCTATTTGTTTCTTCTGCAATTTACATGATTTGATGTCGCGTACCA 2152
5503 CAGGATCTATTTGTTTCTTCTGCAATTTACATGATTTGATGTCGCGTACCA 5444
2153 CCACCCCCCAGTGGGGGGGTGAGAGGGGTGCAAGCCCTGCTCTCACTTTTCTA 2212
5443 CCACCCCCCAGTGGGGGGGTGAGAGGGGTGCAAGCCCTGCTCTCACTTTTCTA 5384
2213 CCTGGAAGTATTTGATTAATCACTTCTGTTTGTAGTTTCA 2260
5383 CCTGGAAGTATTTGATTAATCACTTCTGTTTGTAGTTTCA 5336

RESULT 14
US-09-764-891-10135
Sequence 10135, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
PRIORITY FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0

[illegible]

```
NAME/KEY: SITE
LOCATION: (1342)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1343)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1344)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1345)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1346)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1347)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1348)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1349)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1350)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1351)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1352)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1353)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1354)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1355)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
OTHER INFORMATION: n equals a,t,g, or c
```

Query Match 31.1%; Score 708; DB 11; Length 32174;

Best Local Similarity 100.0%; Pred.No. 0;

Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
1553 GGTGGCTGTGGGAGCTGTGCAAGAACTAGTCTCCCTTTTCAAGCACTCTTTG 1612
26132 GGTGGCTGTGGGAGCTGTGCAAGAACTAGTCTCCCTTTTCAAGCACTCTTTG 26191
1613 CCCCAGAGCAGAGATGAGAAAGCCAGAGAGTGGAGATGATGCTTTCAGCTGTGCT 1672
26192 CCCCAGAGCAGAGATGAGAAAGCCAGAGAGTGGAGATGATGCTTTCAGCTGTGCT 26251
1673 CTGCTGCCAGCAAGTCTTCACTTTGGGCAAGGGGAACTTTTGTGAGAGGGCT 1732
26252 CTGCTGCCAGCAAGTCTTCACTTTGGGCAAGGGGAACTTTTGTGAGAGGGCT 26311
1733 CTGCTTTGTCAAGCAGCGTGAATGAGTGGCGGATCTCAGCTCAGCGCAACTCCAC 1792
26312 CTGCTTTGTCAAGCAGCGTGAATGAGTGGCGGATCTCAGCTCAGCGCAACTCCAC 26371
1793 CTCTGGGTTCAAGTGAATTTCTGCTCAGCTTCCCAAGTGTGGGAAATACAGGCAAG 1852
26372 CTCTGGGTTCAAGTGAATTTCTGCTCAGCTTCCCAAGTGTGGGAAATACAGGCAAG 26431
1853 CCACCATGCCAGCTAATTTTGTATTTTCACTAGAAAGGGATTTCAACGATGGGCA 1912
26432 CCACCATGCCAGCTAATTTTGTATTTTCACTAGAAAGGGATTTCAACGATGGGCA 26491
1913 GGCTGGTCTGGAATCTCTGACGCAAGTATCACCCTCCGCTCCCAAGTGTGG 1972
```

```
26492 GGTGGTCTGGAATCTCTGACCGCAAGTATCACCCTCCGCTCCCAAGTGTCTG 26551
1973 GATTACAGCGTGAAGCCAGCGTCCCGGCCCAAGGGGAACTTTGTGGAGAGCAGA 2032
26552 GATTACAGCGTGAAGCCAGCGTCCCGGCCCAAGGGGAACTTTGTGGAGAGCAGA 26611
2033 GGGGCTCAATCTCCCTCTGATTTCCCATGACATTCGCTTATCTCCCATCTAGC 2092
26612 GGGGCTCAATCTCCCTCTGATTTCCCATGACATTCGCTTATCTCCCATCTAGC 26671
2093 CAGAACTATTGTGTTTCTCTGCAATTTACTATGATTTGATGTGCGGTACCA 2152
26672 CAGAACTATTGTGTTTCTCTGCAATTTACTATGATTTGATGTGCGGTACCA 26731
2153 CCACCCCCCATGAGGGGGGTGAGAGGGGTGCAAGGCCCTGCTCCACTTTTCTA 2212
26732 CCACCCCCCATGAGGGGGGTGAGAGGGGTGCAAGGCCCTGCTCCACTTTTCTA 26791
2213 CCTTGAATCTGATTTGATTAATCACTCTGTTGTGAGTTTCA 2260
26792 CCTTGAATCTGATTTGATTAATCACTCTGTTGTGAGTTTCA 26839
```

RESULT 15

US-09-764-891-10179/c

Sequence 10179, Application US/09764891

Publication No. US20030077808A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC006

CURRENT APPLICATION NUMBER: US/09/764,891

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 10231

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 10179

LENGTH: 32174

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (29356)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29357)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29358)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29359)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29360)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29361)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29362)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29363)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29364)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (29365)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

[illegible]

```

; NAME/KEY: SITE
; LOCATION: (29415)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (29416)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE

```

```

Query Match      31.1%; Score 708; DB 11; Length 32174;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1553 GGTGGCTGTGGGAGTGTGAGAACTACTGTCCTCTTTTCACAGCACTCCTTG 1612
DB 6043 GGTGGCTGTGGGAGTGTGAGAACTACTGTCCTCTTTTCACAGCACTCCTTG 5984
QY 1613 CCCAGAGCAGAGATGAGAAAGCAGAGAGTGAATGATGCTTCCAGCTGTGCT 1672
DB 5983 CCCAGAGCAGAGATGAGAAAGCAGAGAGTGAATGATGCTTCCAGCTGTGCT 5924
QY 1673 CTGCTGCAGCCCAAGTCTTATTTGGGCCAAAGGGGAACTTTTGTGAGAGCGT 1732
DB 5923 CTGCTGCAGCCCAAGTCTTATTTGGGCCAAAGGGGAACTTTTGTGAGAGCGT 5864
QY 1733 CTGCTTTGTACCCAGCTGGAATGCAGTGGCGGATCTCAGCTCACCGCACTCCAC 1792
DB 5863 CTGCTTTGTACCCAGCTGGAATGCAGTGGCGGATCTCAGCTCACCGCACTCCAC 5804
QY 1793 CTCCTGGGTTCAAGTATTTTCCCTGAGCTCCCAAGTACGAGGAAATACAGGACG 1852
DB 5803 CTCCTGGGTTCAAGTATTTTCCCTGAGCTCCCAAGTACGAGGAAATACAGGACG 5744
QY 1853 CCACCATGCCCAGCTAATTTTGTATTTTCAGTAGAAACGGGATTTCAACAGTTGGCCA 1912
DB 5743 CCACCATGCCCAGCTAATTTTGTATTTTCAGTAGAAACGGGATTTCAACAGTTGGCCA 5684
QY 1913 GGTGCTGTGAACTCTGAGACCGCAAGTATCCACCGGCTCCGCTCCCAAGTGTGG 1972
DB 5683 GGTGCTGTGAACTCTGAGACCGCAAGTATCCACCGGCTCCGCTCCCAAGTGTGG 5624
QY 1973 GATTACAGGCGTGAAGCCACGCTGCCGCGCCCAAGGGGAACTCTGTGGAGAGCAGA 2032
DB 5623 GATTACAGGCGTGAAGCCACGCTGCCGCGCCCAAGGGGAACTCTGTGGAGAGCAGA 5564
QY 2033 GGGGCTCAATCTCCCTCTGATTCGCCATGACATTCCTTATCTCTCCCATCTAGC 2092
DB 5563 GGGGCTCAATCTCCCTCTGATTCGCCATGACATTCCTTATCTCTCCCATCTAGC 5504
QY 2093 CAGGAATCTATTTGTTTTTCTTCCCAATTTACTATGATGTGTATGCGCGTACCA 2152
DB 5503 CAGGAATCTATTTGTTTTTCTTCCCAATTTACTATGATGTGTATGCGCGTACCA 5444
QY 2153 CCACCCCCCAGTGGGGGGTGAAGAGGGGCTGCTGCTGCTCACTTTTCTA 2212
DB 5443 CCACCCCCCAGTGGGGGGTGAAGAGGGGCTGCTGCTGCTCACTTTTCTA 5384
QY 2213 CTTTGAATCTATTAATAATCACTTCTGTTTGTCAAGTTTCA 2260
DB 5383 CTTTGAATCTATTAATAATCACTTCTGTTTGTCAAGTTTCA 5336

```

Search completed: November 22, 2003, 14:21:35
Job time : 707 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic : nucleic search, using sw model

Run on: November 22, 2003, 08:15:16 / Search time 135 Seconds
(without alignments)
7431.584 Million cell updates/sec

Title: US-09-938-803-26

Perfect score: 2273

Sequence: 1 99999TGAAGCAGTACGTTT.....ttttcaaaaaaaaaaaaaa 2273

Scoring table: OLIGO_NJC

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/1/ina/5A COMB.seq:*
- 2: /cgn2_6/prodata/1/ina/5B COMB.seq:*
- 3: /cgn2_6/prodata/1/ina/6A COMB.seq:*
- 4: /cgn2_6/prodata/1/ina/6B COMB.seq:*
- 5: /cgn2_6/prodata/1/ina/PCTUS COMB.seq:*
- 6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	2.6	74962	4 US-09-685-853A-3	Sequence 3, Appl1
2	49	2.2	907	4 US-09-996-243-398	Sequence 398, App
3	49	2.2	174493	4 US-09-804-471A-3	Sequence 3, Appl1
4	47	2.1	1252	3 US-09-305-384-7	Sequence 7, Appl1
5	47	2.1	2280	4 US-09-702-705-321	Sequence 321, App
6	47	2.1	3129	4 US-09-736-457-321	Sequence 321, App
7	47	2.1	3129	4 US-10-045-428A-9	Sequence 9, Appl1
8	47	2.1	6235	3 US-09-305-384-5	Sequence 9, Appl1
9	47	2.1	6235	4 US-09-525-160B-6	Sequence 6, Appl1
10	47	2.1	6679	3 US-09-305-384-1	Sequence 1, Appl1
11	47	2.1	6679	4 US-09-525-160B-5	Sequence 1, Appl1
12	47	2.1	7210	2 US-08-257-963B-10	Sequence 10, Appl1
13	47	2.1	7210	4 US-08-367-841A-10	Sequence 10, Appl1
14	47	2.1	7210	5 PCT-US95-07201-10	Sequence 10, Appl1
15	47	2.1	10825	3 US-08-652-265-1	Sequence 1, Appl1
16	47	2.1	10825	3 US-08-652-265-3	Sequence 3, Appl1
17	47	2.1	10825	3 US-08-652-265-5	Sequence 5, Appl1
18	47	2.1	10825	3 US-08-652-265-7	Sequence 7, Appl1
19	47	2.1	10825	3 US-08-834-497A-1	Sequence 1, Appl1
20	47	2.1	10825	3 US-08-834-497A-3	Sequence 3, Appl1
21	47	2.1	10825	3 US-08-834-497A-5	Sequence 5, Appl1
22	47	2.1	10825	3 US-08-834-497A-7	Sequence 7, Appl1
23	47	2.1	10825	3 US-09-503-444A-1	Sequence 1, Appl1
24	47	2.1	10825	3 US-09-503-444A-3	Sequence 3, Appl1
25	47	2.1	10825	3 US-09-503-444A-5	Sequence 5, Appl1
26	47	2.1	10825	3 US-09-503-444A-7	Sequence 7, Appl1
27	47	2.1	12146	4 US-09-277-457-27	Sequence 27, Appl1

28	47	2.1	12146	4 US-09-679-729-27	Sequence 27, Appl1
29	47	2.1	14581	4 US-08-520-372D-4	Sequence 4, Appl1
30	47	2.1	18000	4 US-09-657-346A-17	Sequence 17, Appl1
31	47	2.1	22481	4 US-08-367-841A-43	Sequence 43, Appl1
32	47	2.1	22481	5 PCT-US95-07201-43	Sequence 43, Appl1
33	47	2.1	22484	4 US-09-875-223-2	Sequence 2, Appl1
34	47	2.1	39982	4 US-09-820-924-3	Sequence 3, Appl1
35	47	2.1	99500	4 US-09-798-096-10	Sequence 10, Appl1
36	47	2.1	116592	4 US-09-818-512-3	Sequence 3, Appl1
37	47	2.1	162450	4 US-09-345-882-1	Sequence 1, Appl1
38	47	2.1	246240	2 US-08-724-394A-20	Sequence 20, Appl1
39	47	2.1	246240	2 US-08-724-394A-21	Sequence 21, Appl1
40	47	2.1	246240	2 US-08-724-394A-22	Sequence 22, Appl1
41	46	2.0	289	2 US-08-481-658B-63	Sequence 63, Appl1
42	46	2.0	289	2 US-08-477-504A-63	Sequence 63, Appl1
43	46	2.0	289	2 US-08-486-756A-63	Sequence 63, Appl1
44	46	2.0	289	2 US-08-485-862B-63	Sequence 63, Appl1
45	46	2.0	289	3 US-08-787-739-63	Sequence 63, Appl1

ALIGNMENTS

```

RESULT 1
US-09-685-853A-3
; Sequence 3, Application US/09685853A
; Patent No. 6479270
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000871
; CURRENT APPLICATION NUMBER: US/09/685, 853A
; PRIOR FILING DATE: 2002-05-06
; PRIOR APPLICATION NUMBER: 60/182,194
; PRIOR FILING DATE: 2000-02-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 74962
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc:feature
; LOCATION: (1)...(74962)
; OTHER INFORMATION: n = A,T,C or G
US-09-685-853A-3

Query Match
Best Local Similarity 2.6%; Score 59; DB 4; Length 74962;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1782 GCAACCTCCACCTCTGCTTCAAGTATTTCTCTCAGCTCCAGCTCCAGTACTGGG 1840
Db 62149 GCAACCTCCACCTCTGCTTCAAGTATTTCTCTCAGCTCCAGCTCCAGTACTGGG 62207

RESULT 2
US-09-996-243-398/C
; Sequence 398, Application US/09996243
; Patent No. 6478825
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, Audrey

```

APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kijavijn, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2730P1C13
CURRENT APPLICATION NUMBER: US/09/996,243
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088126
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09

PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535

OTHER INFORMATION: N = A.T.C OF G

US-09-60-50

Query Match 2.1%; Score 47; DB 4; Length 2280;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1956 GCCTCCCAAGTGTGGATTACAGGCGTGAGCCACCGTCCCGGCC 2002
 DB 1113 GCCTCCCAAGTGTGGATTACAGGCGTGAGCCACCGTCCCGGCC 1159

RESULT 6

US-09-736-457-321
 ; Sequence 321, Application US/09736457
 ; Patent No. 6509448
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tonglong
 ; APPLICANT: Bangur, Chaitanya S.
 ; APPLICANT: Lodes, Michael A.
 ; APPLICANT: Fanger, Gary
 ; APPLICANT: Vedvick, Tom
 ; APPLICANT: Carter, Darick
 ; APPLICANT: Retter, Marc
 ; APPLICANT: Mannion, Jane
 ; APPLICANT: Pan, Liqun
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE REFERENCE: 210121.478C15
 ; CURRENT APPLICATION NUMBER: US/09/736,457
 ; NUMBER OF SEQ ID NOS: 1864
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 321
 ; LENGTH: 2280
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-736-457-321

Query Match 2.1%; Score 47; DB 4; Length 2280;
 Best Local Similarity 100.0%; Pred. No. 2.1e-11;
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1956 GCCTCCCAAGTGTGGATTACAGGCGTGAGCCACCGTCCCGGCC 2002
 DB 1113 GCCTCCCAAGTGTGGATTACAGGCGTGAGCCACCGTCCCGGCC 1159

RESULT 7

US-10-045-428A-9
 ; Sequence 9, Application US/10045428A
 ; Patent No. 6500943
 ; GENERAL INFORMATION:
 ; APPLICANT: Mano, Hiroyuki
 ; APPLICANT: Sakata, Tsunaki
 ; APPLICANT: Hasegawa, Mamoru
 ; APPLICANT: Tabata, Toshiaki
 ; TITLE OF INVENTION: Promoter
 ; FILE REFERENCE: 50026/011003
 ; CURRENT APPLICATION NUMBER: US/10/045,428A
 ; CURRENT FILING DATE: 2002-04-15
 ; PRIOR APPLICATION NUMBER: 09/735,103
 ; PRIOR FILING DATE: 2000-12-12
 ; PRIOR APPLICATION NUMBER: 09/142,529
 ; PRIOR FILING DATE: 1999-08-12
 ; PRIOR APPLICATION NUMBER: PCT/JP97/00741
 ; PRIOR FILING DATE: 1997-03-10
 ; PRIOR APPLICATION NUMBER: JP 8/54294
 ; PRIOR FILING DATE: 1996-03-12
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 9
 ; LENGTH: 3129
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-045-428A-9

Query Match 2.1%; Score 47; DB 4; Length 3129;
 Best Local Similarity 100.0%; Pred. No. 2.1e-11;

Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1956 GCCTCCCAAGTGTGGATTACAGGCGTGAGCCACCGTCCCGGCC 2002
 DB 410 GCCTCCCAAGTGTGGATTACAGGCGTGAGCCACCGTCCCGGCC 456

RESULT 8

US-09-305-384-5/c
 ; Sequence 5, Application US/09305384
 ; Patent No. 6242218
 ; GENERAL INFORMATION:
 ; APPLICANT: Treco, Douglas A.
 ; APPLICANT: Heasley, Michael W.
 ; APPLICANT: Seiden, Richard F.
 ; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
 ; FILE REFERENCE: 07236/017001
 ; CURRENT APPLICATION NUMBER: US/09/305,384
 ; CURRENT FILING DATE: 1999-05-05
 ; EARLIER APPLICATION NUMBER: US 60/084,649
 ; EARLIER FILING DATE: 1998-05-07
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 5
 ; LENGTH: 6235
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-305-384-5

Query Match 2.1%; Score 47; DB 3; Length 6235;
 Best Local Similarity 100.0%; Pred. No. 2e-11;
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1956 GCCTCCCAAGTGTGGATTACAGGCGTGAGCCACCGTCCCGGCC 2002
 DB 5251 GCCTCCCAAGTGTGGATTACAGGCGTGAGCCACCGTCCCGGCC 5205

RESULT 9

US-09-525-1608-6/c
 ; Sequence 6, Application US/095251608
 ; Patent No. 6569681
 ; GENERAL INFORMATION:
 ; APPLICANT: Ivanov, Evgenii
 ; TITLE OF INVENTION: METHODS OF IMPROVING HOMOLOGOUS RECOMBINATION
 ; FILE REFERENCE: 10278/016001
 ; CURRENT APPLICATION NUMBER: US/09/525,1608
 ; CURRENT FILING DATE: 2000-03-14
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 6
 ; LENGTH: 6235
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-525-1608-6

Query Match 2.1%; Score 47; DB 4; Length 6235;
 Best Local Similarity 100.0%; Pred. No. 2e-11;
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1956 GCCTCCCAAGTGTGGATTACAGGCGTGAGCCACCGTCCCGGCC 2002
 DB 5251 GCCTCCCAAGTGTGGATTACAGGCGTGAGCCACCGTCCCGGCC 5205

RESULT 10

US-09-305-384-1/c
 ; Sequence 1, Application US/09305384
 ; Patent No. 6242218
 ; GENERAL INFORMATION:
 ; APPLICANT: Treco, Douglas A.
 ; APPLICANT: Heasley, Michael W.
 ; APPLICANT: Seiden, Richard F.

TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY
FILE REFERENCE: 07236/017001
CURRENT APPLICATION NUMBER: US/09/305,384
CURRENT FILING DATE: 1999-05-05
EARLIER APPLICATION NUMBER: US 60/084,649
EARLIER FILING DATE: 1998-05-07
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 6679
TYPE: DNA
ORGANISM: Homo sapiens
US-09-305-384-1

Query Match
Best Local Similarity 100.0%; Pred. No. 2e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1956 GCCTCCCAAGTCTGGATTACAGCGCGTACCGTCCCGGCC 2002
DB 5270 GCCTCCCAAGTCTGGATTACAGCGCGTACCGTCCCGGCC 5224

RESULT 11
US-09-525-160B-5/C
Sequence 5, Application US/09525160B
Patent No. 6569681
GENERAL INFORMATION:
APPLICANT: Ivanov, Evgenii
TITLE OF INVENTION: METHODS OF IMPROVING HOMOLOGOUS RECOMBINATION
FILE REFERENCE: 10278/016001
CURRENT APPLICATION NUMBER: US/09/525,160B
CURRENT FILING DATE: 2000-03-14
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 6679
TYPE: DNA
ORGANISM: Homo sapiens
US-09-525-160B-5

Query Match
Best Local Similarity 100.0%; Pred. No. 2e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1956 GCCTCCCAAGTCTGGATTACAGCGCGTACCGTCCCGGCC 2002
DB 5270 GCCTCCCAAGTCTGGATTACAGCGCGTACCGTCCCGGCC 5224

RESULT 12
US-08-257-963B-10/C
Sequence 10, Application US/08257963B
Patent No. 5840686
GENERAL INFORMATION:
APPLICANT: Chader, Gerald J.; Becerra, S.
APPLICANT: Patricia; Schwartz, Joan P.;
APPLICANT: Taniwaki, Takayuki
TITLE OF INVENTION: PIGMENT EPITHELIUM
TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION OF ITS NOVEL
TITLE OF INVENTION: BIOLOGICAL ACTIVITY AND SEQUENCES ENCODING
TITLE OF INVENTION: AND EXPRESSING THE PROTEIN
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morgan & Finnegan
STREET: 345 Park Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,963B
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/952,796
FILING DATE: 24-SEPT-1992
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36434
REFERENCE/DOCKET NUMBER: 20264126US1
TELEPHONE: (212) 758-4800
FAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 7210 Base Pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Unknown
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Human
IMMEDIATE SOURCE:
LIBRARY: DASH II
FEATURE:
NAME/KEY: JT106
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: 7.2 kb No. 5840686 1 fragments
US-08-257-963B-10

Query Match
Best Local Similarity 100.0%; Pred. No. 2e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1956 GCCTCCCAAGTCTGGATTACAGCGCGTACCGTCCCGGCC 2002
DB 5615 GCCTCCCAAGTCTGGATTACAGCGCGTACCGTCCCGGCC 5569

RESULT 13
US-08-367-841A-10/C
Sequence 10, Application US/08367841A
Patent No. 6319687
GENERAL INFORMATION:
APPLICANT: Chader, Gerald J.; Rodriguez,
APPLICANT: Ignacio R.; Mazuruk, Krzysztof;
APPLICANT: Tomblin-Tink, Joyce
TITLE OF INVENTION: PIGMENT EPITHELIUM
TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morgan & Finnegan
STREET: 345 Park Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/367,841A
FILING DATE: 30-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/257,963
FILING DATE: 07-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/952,796
FILING DATE: 24-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36434
REFERENCE/DOCKET NUMBER: 20264126US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 7210 Base Pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Unknown
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Human
IMMEDIATE SOURCE:
LIBRARY: DASH II
FEATURE:
NAME/KEY: JTEA
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: 7.0 kb No. 6319687 1-No. 6319687
OTHER INFORMATION: fragment; derived from human placental
OTHER INFORMATION: genomic DNA; also referred to as JTI06
US-08-367-841A-10

Query Match 2.1%; Score 47; DB 4; Length 7210;
Best Local Similarity 100.0%; Pred. No. 2e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1956 GCCTCCCAAGTCTGGATTACAGCGGTGACCCAGCGTCCCGGCC 2002
DB 5615 GCCTCCCAAGTCTGGATTACAGCGGTGACCCAGCGTCCCGGCC 5569

RESULT 14
PCT-US95-07201-10/c
Sequence 10, Application PC/TUS9507201
GENERAL INFORMATION:
APPLICANT: Chader, Gerald J.; Becerra, Sofia
APPLICANT: Patricia; Schwartz, Joan P.;
APPLICANT: Tanigaki, Takayuki
TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morgan & Finnegan, L.L.P.
STREET: 345 Park Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07201
FILING DATE: 06-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/367,841
FILING DATE: 30-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/257,963

FILING DATE: 07-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/952,796
FILING DATE: 24-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36434
REFERENCE/DOCKET NUMBER: 20264126PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 7210 Base Pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Unknown
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Human
IMMEDIATE SOURCE:
LIBRARY: DASH II
FEATURE:
NAME/KEY: JTEA
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: 7.0 kb Not 1-Not
OTHER INFORMATION: fragment; Derived from human placental
OTHER INFORMATION: genomic DNA; also referred to as JTI06
PCT-US95-07201-10

Query Match 2.1%; Score 47; DB 5; Length 7210;
Best Local Similarity 100.0%; Pred. No. 2e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1956 GCCTCCCAAGTCTGGATTACAGCGGTGACCCAGCGTCCCGGCC 2002
DB 5615 GCCTCCCAAGTCTGGATTACAGCGGTGACCCAGCGTCCCGGCC 5569

RESULT 15
US-08-652-265-1/c
Sequence 1, Application US/08652265
Patent No. 6025130
GENERAL INFORMATION:
APPLICANT: Thomas, Winston J.
APPLICANT: Drayna, Dennis T.
APPLICANT: Feder, John N.
APPLICANT: Gnlrke, Andreas
APPLICANT: Ruddy, David
APPLICANT: Tsuchihashi, Zenta
TITLE OF INVENTION: Hereditary Hemochromatosis Gene
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/652,265
FILING DATE: 23-MAY-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M.
REGISTRATION NUMBER: 30/223

```

REFERENCE/DOCKET NUMBER: 17957-000500
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (415) 576-0200
  TELEFAX: (415) 576-0300
  INFORMATION FOR SEQ ID NO: 1:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 10825 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
  FEATURE:
    NAME/KEY: CDS
    LOCATION: join(361..436,3762..4025,4235..4510,5606..5881,
    OTHER INFORMATION: /product="Hereditary Hemochromatosis
    OTHER INFORMATION: /note="No. 6025130mal or wild-type (unaffected)
    OTHER INFORMATION: Hereditary Hemochromatosis (HH) gene
  FEATURE:
    NAME/KEY:
    LOCATION: 140..7319
    OTHER INFORMATION: /note="start and stop positions for
    OTHER INFORMATION: normal or wild-type (unaffected) allele
  FEATURE:
    NAME/KEY:
    LOCATION: 3852..3891
    OTHER INFORMATION: /note="start and stop positions for
    OTHER INFORMATION: normal or wild-type (unaffected) genomic
    OTHER INFORMATION: sequence surrounding variant for 24d2 (C)
  FEATURE:
    NAME/KEY:
    LOCATION: 5507..6023
    OTHER INFORMATION: /note="start and stop positions for
    OTHER INFORMATION: normal or wild-type (unaffected) genomic
    OTHER INFORMATION: sequence surrounding variant for 24d1 (G)
  FEATURE:
    NAME/KEY: allele
    LOCATION: replace(3872,"c")
    OTHER INFORMATION: /phenotype="normal or wild-type
    OTHER INFORMATION: (unaffected)"
  FEATURE:
    NAME/KEY: allele
    LOCATION: replace(3878,"a")
    OTHER INFORMATION: /phenotype="normal or wild-type
    OTHER INFORMATION: (unaffected)"
  FEATURE:
    NAME/KEY: allele
    LOCATION: replace(5834,"g")
    OTHER INFORMATION: /phenotype="normal or wild-type
    OTHER INFORMATION: (unaffected)"
  OTHER INFORMATION: /label=24d1
US-08-652-265-1

Query Match      2.1%; Score 47; DB 3; Length 10825;
Best Local Similarity 100.0%; Pred. No. 2e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1956 GCCTCCCAAGTGCCTGGATTACAGCGGTGAGCCACCGTCCCGGCC 2002
DB 8175 GCCTCCCAAGTGCCTGGATTACAGCGGTGAGCCACCGTCCCGGCC 8129

```

Search completed: November 22, 2003, 14:01:17
 Job time: 137 secs

THIS PAGE BLANK (USPTO)